

Reading comprehension and listening comprehension in children: An individual differences investigation

Sian Roberts

A thesis submitted in partial fulfilment for the requirements of the degree
of Doctor of Philosophy

Sponsoring Establishment
University of Central Lancashire
Preston,
PR1 2HE.

Submitted September 2010



Student Declaration

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution.

I declare that no material contained in the thesis has been used in any other submission for an academic award and that it is solely my own work.

Signature of Candidate _____

Type of Award _____

School _____

ABSTRACT

Little research has explored listening comprehension in children whereas reading comprehension has been extensively investigated. One of the reasons for this is that listening comprehension and reading comprehension are highly correlated and it is generally assumed that they draw on the same cognitive-linguistic processes. This assumption has been formalised in the “Simple View of Reading” (Gough & Tunmer, 1986) which states that, once printed text has been decoded, it is understood in exactly the same way as its spoken equivalent. The main aim of the work presented in this thesis was to investigate the assumption that the same skills and processes underpin reading comprehension and listening comprehension by conducting an investigation of the demands made by comprehension in each modality which are over and above those shared with comprehension in the other modality. This issue has not previously been addressed.

Children were assessed on both standardised and true/false measures of listening comprehension and reading comprehension and on several variables previously found to predict reading comprehension. Although results varied slightly according to the measure of comprehension used, broad support was found for the Simple View of Reading as a conceptual framework for explaining reading comprehension. It appeared, however, that listening comprehension involved skills which were not shared with reading comprehension. Of particular interest was the finding that, compared to reading comprehension, listening comprehension appeared to make extra demands on children’s inferencing ability. In a further study it was ascertained that this was not simply due to the shared memory demands of the inferencing and listening comprehension tasks. The hypothesis that listening comprehension ability depends on the ability to generate inferences “on-line” whilst listening was tested in a final study but was not supported.

In conclusion, the research presented here suggests that listening comprehension is a topic worthy of investigation in its own right and that, for purposes of both research and educational practice, children’s comprehension is best assessed in both modalities.

CONTENTS

Chapter 1.	1
Overview of Thesis	1
Chapter 2.	6
Discourse comprehension and the Simple View of Reading	6
2.1. Introduction	6
2.2. Discourse comprehension	6
2.3. The Construction-Integration (CI) model of comprehension	8
2.4. Reading comprehension vs listening comprehension	10
2.5. The Simple View of Reading	12
2.5.1. Description	12
2.5.2. Supporting evidence	14
2.5.3. Limitations	19
2.5.4. Implications	27
2.5.5. Rationale for current research	27
Chapter 3.	30
Components of comprehension	30
3.1. Introduction	30
3.2. Word-level skills	31
3.2.1. Word-recognition/ decoding ability	31
3.2.2. Vocabulary and semantic knowledge	32
3.3. Syntactic skills	36
3.4. Inferencing	38
3.5. Exposure to print	46
3.6. Working Memory	48
3.7. Non-Verbal Intelligence	53
3.8. Summary and introduction to Study 1	55
Chapter 4.	58
Study 1: Are listening comprehension and reading comprehension underpinned by the same cognitive-linguistic skills? An investigation using standardised tests.	58
4.1. Summary of aims	58
4.2. Method	58

4.3. Results.....	71
4.4. Discussion.....	87
Chapter 5.	93
Study 2: Are listening comprehension and reading comprehension underpinned by the same cognitive-linguistic skills? An investigation using true/false tests....	93
5.1. Introduction.....	93
5.2. Summary of aims	96
5.3. Method	96
5.4. Results.....	99
5.5. Discussion.....	112
5.6. Summary of Studies 1 and 2	115
Chapter 6.	117
Study 3: Memory for explicit information and inference generation in listening comprehension.....	117
6.1. Introduction.....	117
6.2. Summary of aims	121
6.3. Method	122
6.4. Results.....	126
6.5. Discussion.....	132
6.6. Summary of Study 3	134
Chapter 7.	135
Study 4: On-line inferencing and listening comprehension	135
7.1. Introduction.....	135
7.2. Summary of Aims	140
7.3. Method	141
7.4. Results.....	147
7.5. Follow-up study	154
7.6. Discussion.....	157
Chapter 8.	161
General Discussion.....	161
8.1. Background to studies.....	161
8.2. Summary of findings.....	162
8.3. Implications for theory and research.....	169
8.4. Practical implications.....	171

8.5. Limitations and Future Research	172
8.6. Summary/ Original contributions	174
References.....	176
Appendix 1: Title Recognition Test.....	200
Appendix 2: Mean Responses to TOWK Items	201
Appendix 3: Off-line inferencing test material	202
Appendix 4: True/False Listening Comprehension Test.....	204
Appendix 5: True/False Reading Comprehension Test	207
Appendix 6: Off-line inferencing and memory questions	210
Appendix 7: On-line inferencing stimulus materials.....	212

LIST OF TABLES

Table 1: Example story from the test of inferencing ability	66
Table 2: Means and standard deviations for all Study 1 measures (raw scores)	74
Table 3: Correlations between Study 1 measures	75
Table 4: Standard regression analysis with NARA-II reading comprehension as DV and language, memory and intelligence variables as predictors.....	76
Table 5: Standard regression analysis with LCTS listening comprehension as DV and language, memory and intelligence variables as predictors.....	77
Table 6: Summary of standard regression analyses showing beta coefficients for variables predicting unique and significant variance in comprehension in each modality	78
Table 7: Standard regression analysis with NARA-II reading comprehension residuals as DV and language, memory and intelligence variables as predictors.....	81
Table 8: Standard regression analysis with LCTS listening comprehension residuals as DV and language, memory and intelligence variables as predictors	82
Table 9: Summary of standard regression analyses showing beta coefficients for variables predicting unique and significant variance in residualised measures of comprehension in each modality	83
Table 10: Example story from true/false test of listening comprehension	98
Table 11: Means and standard deviations for d-prime scores on forced-choice measures of comprehension.....	102
Table 12: Bivariate correlations between measures of comprehension.....	102
Table 13: Correlations between Study 2 measures	104
Table 14: Standard regression analysis with d-prime scores on true/false reading comprehension as DV and language, memory and intelligence variables as predictors	105
Table 15: Standard regression analysis with d-prime scores on true/false listening comprehension as DV and language, memory and intelligence variables as predictors	106
Table 16: Summary of standard regression analyses showing beta coefficients for variables predicting unique and significant variance in standardised and in true/false measures of comprehension.....	107

Table 17: Standard regression analysis with true/false reading comprehension residuals as DV and language, memory and intelligence variables as predictors	109
Table 18: Standard regression analysis with true/false listening comprehension residuals as DV and language, memory and intelligence variables as predictors	110
Table 19: Summary of standard regression analyses showing beta coefficients for variables predicting unique and significant variance in residualised standardised and true/false measures of comprehension	111
Table 20: Example story from the test of inferencing ability and memory for explicit information.....	125
Table 21: Means and standard deviations for all Study 3 measures (raw scores)	128
Table 22: Correlations between Study 3 measures	129
Table 23: Hierarchical regression analyses with LCTS listening comprehension as the criterion variable	130
Table 24: Illustrative examples of the four conditions in which target words appeared	144
Table 25: Means and SD's for response latencies to words in each condition.....	149
Table 26: Means and SD's for response latencies to words in each condition for each comprehension group.....	152
Table 27: Means and SD's for response latencies to words in each condition in follow-up study	156

LIST OF FIGURES

Figure 1: The Simple View of Reading.....	14
Figure 2: Relationship between residualised measures of reading comprehension and listening comprehension, with distribution of children with good and poor word recognition skills shown.....	85
Figure 3: Relationship between residualised measures of reading comprehension and listening comprehension, with distribution of children with good and poor inferencing skills shown.....	86
Figure 4: Nature of stimuli used in on-line inferencing test.....	142
Figure 5: Experimental timeline.....	146
Figure 6: Mean response latencies to words as a function of association and sentence coherence.....	150

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors, John Fisk, Debra Long and Alice Spooner for their contributions to this work. Particular thanks have to go to John and Debra for their encouragement and guidance over the last couple of years.

Many children were involved in my research and I would like to thank them all for their co-operation and enthusiasm. Also, I would like to thank their class teachers and head teachers. Without their support, none of this would have been possible.

I would like to thank my friends, both at UCLan and at home, who have listened to me when it's all felt too much.

Finally, I would like to thank my children, Thomas, Matthew and Rebecca, and my partner, David, who all know more about comprehension than they ever wanted to.

Chapter 1.

Overview of Thesis

The study of discourse comprehension aims to discover how extended texts, presented either in spoken or in written language, are understood. Most research, however, with both children and adults has focused on the comprehension of written texts. There are historical and theoretical reasons why this may be the case.

Historically, although children's reading has been the focus of research interest for many years, the main concern has been the development of word-reading skills (e.g. Goswami & Bryant, 1990). However, it has recently become apparent that approximately 10% of the school population can be classed as poor comprehenders, children who struggle to comprehend written texts despite having competent word reading abilities (Nation & Snowling, 1997; Yuill & Oakhill, 1991). Thus, research into reading comprehension in children has emerged from a wider concern with children's reading development and much is now known about the component skills underpinning reading comprehension success (Cain & Oakhill, 2003).

From a theoretical point of view, the finding that reading comprehension and listening comprehension tend to show moderate to high correlations (Stothard & Hulme, 1992), has led to the assumption that the same cognitive-linguistic processes underpin comprehension in the two modalities. Indeed, Gernsbacher, Varner, and Faust (1990), have postulated the existence of a General Comprehension Skill common to comprehension of written, spoken and pictorial information. However, not all authors agree that reading comprehension and listening comprehension involve exactly the same processes (Rubin, 1980; Carlisle & Felbinger, 1991). The two tasks make different demands on comprehenders and it has been argued that this results in different processing strategies being important in comprehension in each modality (Danks & End, 1987).

The assumption that reading comprehension and listening comprehension are underpinned by the same cognitive-linguistic processes has important practical

implications. It has been formalised in the Simple View of Reading (Gough & Tunmer, 1986) which suggests that once printed words have been decoded, the meaning of the discourse is constructed in exactly the same way as if the words had been listened to. This view has now been integrated into the National Curriculum for England and Wales. The Primary Framework for Literacy encourages teachers to teach word recognition skills and listening comprehension skills in order to ensure reading comprehension success (DfES, 2006).

Placing comprehension skills at the heart of children's literacy education is a welcome development. However, it is important to acknowledge that the current evidence base regarding comprehension is based almost exclusively on research into reading comprehension. Yet, for young children at the beginning stages of learning to read, or older children who struggle with the demands of word recognition, comprehension skills will be taught through the spoken language not the printed text. Furthermore, as Lehto and Anttila (2003) argue, children spend much of their school day engaged in listening activities, no matter how proficient their reading comprehension. For reasons outlined above, however, the evidence base specific to listening comprehension is limited.

The work presented in this thesis was conducted with the aim of exploring the assumption that reading comprehension and listening comprehension are underpinned by the same cognitive–linguistic skills and processes. If this assumption was found to be correct, it could be concluded that findings from research investigating comprehension in one modality also apply to comprehension in the other modality. If, however, support for this assumption was not found, the need to establish separate evidence bases for comprehension in each modality would be highlighted.

The structure of the thesis is as follows. In Chapters 2 and 3, the relevant literature is reviewed. In Chapters 4 – 7, the empirical work undertaken is described. In Chapter 8, findings from the studies are synthesised and critically analysed. The content of each chapter is summarised below.

In Chapter 2, the topic of discourse comprehension is introduced and the influential Construction-Integration Model of comprehension (Kintsch, 1998) is described. A

brief explanation is given of the “unitary view” of comprehension, the assumption that comprehension processes are the same, whether text is written or spoken. Most of the chapter, however, is devoted to an exploration of The Simple View of Reading. A detailed description of this account of reading is given and supporting evidence is reviewed. Some empirical findings have challenged the Simple View of Reading and these are also discussed. Implications of the Simple View of Reading for both research and practice are outlined. In the final section of this chapter, it is argued that an investigation is timely of the assumption that listening comprehension and reading comprehension involve the same cognitive-linguistic skills.

Chapter 3 is a review of the evidence concerning the component skills involved in reading comprehension and listening comprehension. As reading comprehension has been more widely researched than listening comprehension, most of this chapter is devoted to evidence regarding the comprehension of written text. Components covered are word recognition skill, vocabulary and semantic skills, syntactic skills, inferencing ability, exposure to print, working memory and non-verbal intelligence.

Chapter 4 reports the findings from Study 1, a large exploratory study which used regression methods to identify the most important predictors of performance on standardised tests of comprehension in the two modalities. The language, memory and intelligence skills described in Chapter 3 were assessed. It was found that, whilst comprehension in both modalities was predicted by receptive vocabulary skills, reading comprehension appeared to be uniquely predicted by word- and sentence-level language skills whilst listening comprehension was predicted by inferencing and general cognitive skills. Furthermore, the Simple View of Reading was tested directly by identifying the most important predictors of comprehension in each modality having controlled for comprehension in the other modality. As predicted by the Simple View of Reading, the comprehension of written texts was predicted by word recognition skills only once listening comprehension was controlled for. Listening comprehension, however, continued to be predicted by vocabulary and inferencing skills, having controlled for reading comprehension, suggesting that listening comprehension makes additional demands on comprehenders not shared with reading comprehension.

Study 2 was undertaken to address concerns that the results obtained in Study 1 may have arisen from the fact that the standardised tests of reading comprehension and listening comprehension involved different materials and made different demands on comprehenders. Chapter 5 reports the results of Study 2 which was comparable to Study 1 except that true/false tests, which made similar demands on comprehenders, were used to assess comprehension in the two modalities. Again reading comprehension depended on word- and sentence-level language skills whilst listening comprehension was more dependent on higher level cognitive skills, but not, in this case, inferencing. Reading comprehension was predicted by word recognition skills and syntactic skills once listening comprehension was controlled for, providing evidence for a modified version of the Simple View of Reading. Listening comprehension continued to be predicted by non-verbal intelligence, having controlled for reading comprehension, again suggesting that listening comprehension makes demands on comprehenders not shared with reading comprehension. Possible reasons for the differences in the results obtained in Studies 1 and 2 are discussed.

Study 3, reported in Chapter 6, was conducted to further investigate the Study 1 finding that inferencing is an important predictor of performance on the standardised test of listening comprehension. A possible explanation for this finding was that it resulted solely from the shared memory demands of the inferencing task and the comprehension task. In Study 3, this was found not to be the case. Performance on the listening comprehension task was found to be predicted by the ability to make knowledge-based inferences even when memory for the literal content of the text was controlled for. In the conclusion to this chapter, it is speculated that children who engage in more knowledge-based inferencing whilst listening form more coherent and elaborated representations of the text which in turn aids their ability to remember explicitly stated information.

Chapter 7 reports the results from Study 4 which directly addressed the question of whether or not performance on a standardised test of listening comprehension is related to the ability to generate knowledge-based inferences “on-line” whilst listening. Although there was evidence that children appeared to generate inferences as they listened to text, this ability was not found to be related to comprehension ability. Possible reasons for this finding are discussed.

In Chapter 8, results are discussed in detail as are implications for theory, research and practice. It is argued that, although there is evidence that reading comprehension is well explained by the Simple View of Reading, or by a modified version of this conceptual framework, it appears that listening comprehension makes demands on comprehenders over and above those shared with reading comprehension. For the purposes of both research and educational practice, therefore, it appears that comprehension is best assessed in both modalities.

Chapter 2.

Discourse comprehension and the Simple View of Reading

2.1. Introduction

This chapter begins with a brief overview of current understandings of what is meant by discourse comprehension. An influential model of comprehension, the Construction-Integration model (van Dijk & Kintsch, 1983), is then described in some detail. The model is believed to apply to both reading comprehension and listening comprehension and evidence supporting this “unitary view” of comprehension is outlined. Most of this chapter, however, is devoted to exploring the Simple View of Reading (Gough & Tunmer, 1986), a conceptual framework for understanding reading comprehension in terms of the two components of listening comprehension and word recognition. Following a description of the main features of the Simple View of Reading, evidence both supporting and challenging this conceptual framework is reviewed and implications for educational practitioners are discussed. In the concluding section of this chapter, the rationale for conducting the work contained in this thesis is outlined.

2.2. Discourse comprehension

It is currently accepted that the comprehension of written or spoken discourse involves the representation of the situation being described. However, until the 1970's, comprehension was widely regarded as the construction of a representation of the text itself (Zwaan & Radvansky, 1998). All the information necessary to form a semantic representation of a sentence was believed to be contained in the sentence (Spiro, 1980).

Meaning, however, was shown not to exist simply “in the text” in a series of experiments conducted in the 1970's which showed that comprehension is a constructive process which necessarily involves going beyond what is actually stated. In their influential study, Bransford, Barclay, and Franks (1972) showed that the nature of the situation being described affects memory for the text. Sentences such as (1) and (2) below were presented to participants.

1. Three turtles rested *beside* a floating log, and a fish swam beneath *them*.
2. Three turtles rested *on* a floating log, and a fish swam beneath *them*.

Participants were then given recognition tests which included foils such as (3) and (4).

3. Three turtles rested *beside* a floating log, and a fish swam beneath *it*.
4. Three turtles rested *on* a floating log, and a fish swam beneath *it*

It was found that those participants hearing sentence (2) incorrectly “recognised” sentence (4) whereas those hearing sentence (1) did not “false alarm” to sentence (3). The only difference in surface structure between (2) and (4) is the replacement of *them* with *it*. This is also the only difference between (1) and (3). Therefore the incorrect recognition of (4) but not (3) cannot be due to aspects of the surface structure of the text. Bransford et al. concluded that what had been stored was not the surface linguistic information itself but a representation of the situation that was being described by that information. Sentences (2) and (4) describe the same situation whereas sentences (1) and (3) do not.

In 1983, two theories of text comprehension were published which highlighted the importance of the mental representations of the situation described by a text (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). In the first, this representation was known as a “mental model”, in the second as a “situation model”. The theories suggested that a situation or mental model was formed alongside a representation of the text itself. The Construction-Integration model suggested by van Dijk and Kintsch and later developed by Kintsch has been particularly influential (Kintsch, 1988, 1998; Kintsch & Rawson, 2005). It describes the types of information represented in comprehension and the processes involved. A brief overview of the model is set out here.

2.3. The Construction-Integration (CI) model of comprehension

Kintsch and Rawson (2005) argue that text comprehension involves processes at different levels. At the word level, the comprehender has to process individual words and access their meanings. When reading, this also involves the decoding of printed symbols. At the sentence level, word meanings are combined to form idea units or propositions, structures which specify the syntactic and semantic relationships between the words in the sentence. At the text level, propositions are connected to form a coherent microstructure, often using simple inferencing processes such as pronoun identification. Sections of the text are also interrelated, so the connected propositions may themselves be organised into higher order units of meaning called the macrostructure. Together the microstructure and macrostructure comprise the textbase, a representation of the explicit ideas in the text, the information that is actually given.

The textbase representation formed is not sufficient for deep understanding. This requires the construction of a situation model, a mental model of the situation that the text describes. Knowledge-based inferencing is a crucial process in forming a situation model. For example, consider the following passage, taken from Virtue, Parrish, and Jung-Beeman (2008):

From the gate, Walter could see his grandmother coming towards him. After she walked away, he knew that his cheeks would be sore for days.

To understand this passage, comprehenders need to generate the inference, using their general knowledge, that Walter's grandmother must have pinched his cheeks. If this inference is not made, the passage will lack coherence and will not be properly understood.

Kintsch (1998) describes how the CI model explains the integration of information from the text with the comprehender's prior knowledge, to form a representation of the situation described by the text. Prior knowledge includes memories, beliefs, emotions, images and goals. Given idea units in the form of propositions from the text, associated elements from the comprehender's long term memory are retrieved to

form an interrelated network. This retrieval is regarded as primarily a bottom-up process, the result of spreading activation, and the emerging network contains both relevant and irrelevant items. This “construction” process is followed by an “integration” process in which the pattern of activation stabilizes. Those elements that fit together are selectively activated whilst the others are deactivated. Thus, this integration process is one of constraint-satisfaction. Usually, therefore, inferencing is an automatic and spontaneous process. Although the end result, the situation model, is conscious, the processes leading to it are not.

There are, however, many occasions when inferencing is strategic, active, controlled and effortful. Kintsch and Rawson (2005) argue that the importance of this type of inferencing should not be underestimated. However, according to the CI model, it is only when normal comprehension processes break down, and a comprehender fails to understand, that problem solving processes are invoked.

The CI model assumes that the information processing which takes place during comprehension occurs within working memory (Kintsch & Rawson, 2005). For example, if propositions do not co-occur in working memory, connections between them cannot be formed. Similarly, if relevant general knowledge held in long-term memory does not co-occur with textbase propositions, inferences will not be generated. Ericsson and Kintsch (1995) have, however, argued that it is not the capacity of working memory that determines comprehension, but the efficiency with which working memory can be used to encode information from the text and retrieve information from long-term memory.

The CI model highlights the complexity of the language comprehension process and the types of information that need to be represented. It also identifies the component processes involved at word, sentence and discourse level, the role of working memory and the importance of comprehender knowledge. Although it is a model of fluent adult comprehension, much research suggests that the component processes of comprehension which it identifies are involved in children’s discourse comprehension. Empirical evidence concerning these component processes will be reviewed in the next chapter. First, however, the relationship between reading comprehension and listening comprehension will be considered.

2.4. Reading comprehension vs listening comprehension

Introduction

Kintsch and van Dijk (1978) claim that the CI model applies to both listening comprehension and reading comprehension. The assumption that the same processes are involved in comprehension whether the text is presented in print or in speech is sometimes referred to as the “unitary view” of comprehension processes (Horowitz & Samuels, 1987). It is an old concept which pre-dates the formulation of the CI model. For example, Huey (1908, cited in Sticht, 1972) believed reading comprehension to be parasitic upon listening comprehension:

“The child comes to his first reader with the habits of spoken language fairly well formed and these habits grow more deeply set with every year. His meanings inhere in this spoken language and belong but secondarily to the printed symbols”

(Huey, 1908, p.123, cited in Sticht, 1972, p.286)

This “unitary view” of comprehension received empirical support from early studies of adult comprehenders. Sticht (1972) found listening comprehension and reading comprehension for the same material to be equivalent in adult males with poor literacy skills, although many of the participants claimed to find listening comprehension easier. Kintsch and Kozminsky (1977) asked college students to summarise stories that had either been read or heard. They found the content of the summaries to be very similar, regardless of the modality in which the stories had been presented and despite the fact that, in the reading condition, participants were allowed to look back at the text to help them. Kintsch and Kozminsky concluded that reading and listening involve the same comprehension skills.

Other studies of adult comprehension report high correlations between listening comprehension and reading comprehension. Amongst college students, Palmer, MacLeod, Hunt, and Davidson (1985) found correlations as high as .80 between listening comprehension and reading comprehension, whilst Gernsbacher, Varner, and Faust (1990) found a correlation of .92 between the comprehension of written and spoken stories. Furthermore, Gernsbacher et al. also found that the ability to comprehend written and auditory stories was highly correlated with the ability to

comprehend non-verbal picture stories ($r = .82$ and $r = .72$ respectively). Based on this evidence, Gernsbacher et al. proposed the existence of a General Comprehension Skill underpinning comprehension in all modalities.

Empirical support for the “unitary view” has also been found in research involving children. Smiley, Oakley, Worthen, Campione, and Brown (1977) found that, as expected, poor readers aged 13 recalled fewer of the main ideas than good readers when asked to retell stories they had read. The same pattern was also found, however, when the poor and good readers retold stories that they had heard. Similarly, Townsend, Carrithers, and Bever (1987) found that school aged average readers made more errors than skilled readers on comprehension questions following not only stories that they read but also stories that they heard. Furthermore, Berger and Perfetti (1977) found 10-year-old good and poor readers, matched on non-verbal ability, differed in their ability to recall and recognise simple stories whether the stories were read or spoken.

Thus, the evidence reviewed here suggests that reading comprehension and listening comprehension involve the same processes. However, unlike listening comprehension, reading comprehension obviously also involves reading ability. Reading comprehension will not occur if the printed words are not recognised. Only then can processes enabling understanding of the message be employed. It would seem that neither word recognition nor language comprehension skills are by themselves sufficient (Gough & Tunmer, 1986). The dependence of reading comprehension on both these skills may seem particularly important for children who have not yet developed fluent word reading skills and has been formalised in the Simple View of Reading (Gough & Tunmer 1986). This has been enormously influential in affecting both the direction of research and educational policy as will be discussed later. A description of the Simple View of Reading and a review of the research studies that have been conducted to investigate it follow below.

2.5. The Simple View of Reading

2.5.1. Description

Gough and Tunmer (1986) argue that decoding skill and linguistic comprehension make independent contributions to reading comprehension; both are necessary but neither is sufficient. They express this relationship in the formula:

$$R = D \times C,$$

where

R = reading comprehension

D = decoding ability

C = general linguistic comprehension

Decoding (D) is defined as the process of converting print to speech. Hoover and Gough (1990) argue that when children are at the early stages of learning to read, this construct is best assessed using a test of non-word reading which measures children's ability to convert print to sound using phonological information. However, as reading skill improves, D is best measured using a test of real word reading. This distinction is not trivial. As will be seen, the measure of D chosen can affect research findings. In the following discussion "decoding" will refer to the ability to read non-words whilst "word recognition ability" will refer to the ability to read real words. Linguistic comprehension (C) is seen as the process by which discourses are interpreted once word information has been accessed. Critically, this linguistic comprehension process is regarded as the same whether the word is heard or decoded from print:

"the simple view clearly asserts that reading ability should be predictable from a measure of decoding ability...and a measure of listening comprehension."

(Gough & Tunmer, 1986, p.7)

"The simple view presumes that, once the printed matter is decoded, the reader applies to the text exactly the same mechanisms which he or she would bring to bear on its spoken equivalent."

(Gough & Tunmer, 1986, p.9)

Each variable can range from 0 to 1, a score of 1 being an indication of perfection. Gough and Tunmer (1986) emphasise the multiplicative nature of the combination of D and C because if either is zero, reading comprehension will also be zero, regardless of the score on the other variable. The Simple View of Reading implies that decoding and listening comprehension are largely independent of one another.

The Simple View of Reading can be represented schematically as shown in Figure 1 (taken from Stuart & Stainthorp, 2006). Skills vary from good to poor in each of the two domains. Individuals can vary in ability across the domains, and their reading behaviour can be categorised into one of four types depending on the quadrant in which they fall. Children with poor word recognition ability may also have poor language comprehension skills. These children are generally poor readers and are sometimes referred to in the literature as “garden variety poor readers” (Gough & Tunmer, 1986). Other children with poor reading skills, however, do not have a general impairment in language comprehension skills. These children will have impaired reading comprehension ability because of their failure to read words accurately but their listening comprehension skills will be intact. These children with specific problems with word reading exhibit a dyslexic profile. Other children with impaired reading comprehension skills will be those with adequate word recognition skills but a specific deficit in language comprehension processes. These children are referred to as “poor comprehenders”. According to the Simple View of Reading, their comprehension problems will not be specific to reading but will reflect a general problem with language comprehension. Obviously, children falling in the final quadrant will be good reading comprehenders who have both adequate word reading skill and language comprehension ability.

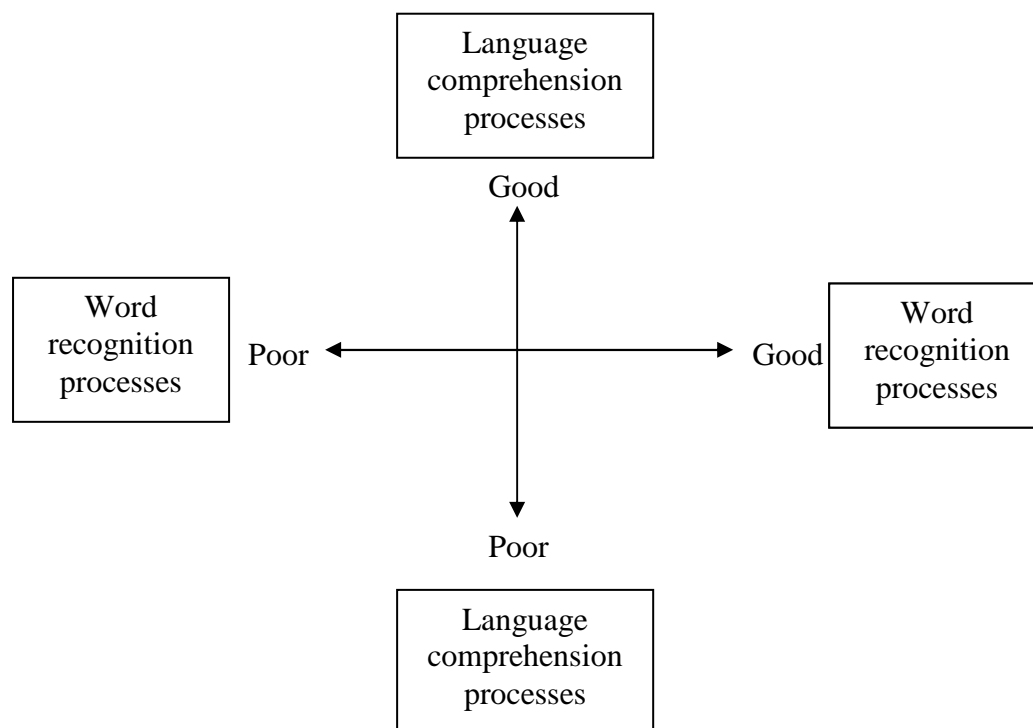


Figure 1: The Simple View of Reading

The Simple View of Reading makes intuitive sense. It is a parsimonious account of reading comprehension, explaining individual differences in comprehension of written material in terms of individual differences in word recognition and listening comprehension skills. Its authors do not deny that reading comprehension is a complex process, but believe that the complexities can be divided into two basic components. According to this account, all the word-, sentence- and discourse-level skills involved in reading comprehension are regarded as subskills of one of these two main components.

2.5.2. Supporting evidence

There is much empirical support for the Simple View of Reading. Research suggests that approximately 10% of children exhibit deficits in reading comprehension despite having adequate word recognition skills (Oakhill, 1982; Yuill & Oakhill, 1991). These poor reading comprehenders perform worse on tests of listening comprehension than good comprehenders, children with equivalent word recognition skills but better reading comprehension scores (Nation & Snowling 1997). Furthermore, whilst the

listening comprehension ability of poor reading comprehenders is worse than that of age-matched good comprehenders, it does not differ from that of younger children matched for reading comprehension (Stothard & Hulme, 1992). The idea that reading comprehension is constrained by listening comprehension ability is further supported by evidence showing that hyperlexic children, children with word recognition skills much better than expected for their age, have a reading comprehension age equivalent to their listening comprehension age (Healy, 1982). Similarly, it has been found that the reading comprehension of children with Down's syndrome is predicted by their listening comprehension skills but not by their word recognition ability which tends to be relatively less impaired (Roch & Levorato, 2009).

Research has demonstrated a double dissociation between children with good word recognition skills but poor listening comprehension skills, and children with poor word recognition skills but good listening comprehension skills. Catts, Hogan, and Fey (2003) found that they were able to classify a large sample of poor readers aged 8 into subgroups based on their relative strengths and weaknesses in word recognition and listening comprehension. These included a poor comprehender group, identified as having adequate word recognition but poor listening comprehension skills, and a poor decoder group with the opposite profile. Furthermore, it was found that those children who were classed as poor comprehenders at age 8 had achieved less well than the poor decoder group on a test of listening comprehension administered at age 6, and subsequently performed less well on a test of listening comprehension at age 10. The poor decoder group also exhibited consistent deficits from the ages of 6 to 10, suggesting that individual differences in word recognition and listening comprehension were stable over time. Shankweiler et al. (1999) similarly reported that 7-9 year old children with high reading comprehension ability relative to their decoding skill differed on measures of listening comprehension from children with low reading comprehension ability relative to decoding skill.

A dissociation between the predictors of word recognition ability and reading comprehension has itself been found. Oakhill, Cain, and Bryant (2003) showed that different cognitive skills explained variance in word recognition and in text comprehension. Reading comprehension, with word recognition ability partialled out, was predicted by integration, metacognitive monitoring and working memory after

statistically controlling for age, non-verbal IQ, verbal IQ and vocabulary. After controlling for the same variables, word recognition ability was predicted by performance on a phonological awareness task. It would be expected, therefore, that children who fall into the “poor comprehender” category would have deficits in higher level language processing skills but intact phonological awareness, whilst poor decoders would have the opposite pattern of impairments. In line with this expectation, Catts, Adlof, and Weismer (2006) found that poor comprehenders aged 14 performed worse than poor decoders on measures of receptive vocabulary and inferencing skill whilst the poor decoders performed worse than poor comprehenders on measures of phonological awareness.

Further evidence of a dissociation between decoding and comprehension skill comes from the factor analysis of datasets which include different measures of reading and language skills. Nation and Snowling (1997) assessed a large sample of 7-10 year-olds using tests of non-word reading, single word reading, word reading in context, reading comprehension and listening comprehension. Whilst word reading and non-word reading loaded heavily on one factor, listening comprehension loaded heavily on another. Reading comprehension loaded onto both factors reflecting the importance of both sets of skills to reading comprehension success. More recently, Kendeou, Savage, and van den Broek (2009) have reported the results of factor analysis on datasets obtained from two studies. In the first study, involving children aged 4 and 6, phonological awareness, letter identification, word identification and vocabulary were found to load on one factor, whilst listening comprehension and comprehension of material presented on a television loaded on another. In the second study, which involved 6-year-old children, it was found that non-word reading fluency, word reading fluency and vocabulary all loaded on to one factor, whilst listening comprehension loaded onto another. As in the Nation and Snowling (1997) study, reading comprehension loaded onto both factors.

Similarly, Structural Equation Modelling has been used to show that, in children aged 4-8, oral language skills, including listening comprehension and the comprehension of material presented on a television, and decoding skills, including letter identification and phonological awareness, form distinct clusters which show continuity over time (Kendeou, van den Broek, White, & Lynch, 2009). This study also reports that, at age

8, children's reading comprehension is found to be independently predicted by these two sets of skills.

The finding that listening comprehension and a measure of decoding or word recognition ability make independent contributions to reading comprehension in children has also been demonstrated in studies using regression techniques. For example, Savage (2001) found that listening comprehension predicted reading comprehension after statistically controlling for either non-word reading or word-recognition in a small sample of teenagers, whilst Byrne and Fielding-Barnsley (1995) showed that listening comprehension, non-word reading and irregular word reading all contributed independently to reading comprehension in 8-year-old children. Nation and Snowling (2004) found that for students aged 8.5, listening comprehension of stories predicted reading comprehension after controlling for nonverbal ability, nonword reading and phonological skills. Furthermore, when the children were 13, the earlier measure of listening comprehension continued to predict reading comprehension even after controlling for the autoregressive effect of earlier reading comprehension. In other words, listening comprehension predicted reading comprehension longitudinally as well as concurrently. The ability of listening comprehension to predict reading comprehension longitudinally has also been reported by de Jong and van der Leij (2002) who showed that the reading comprehension of 9-year-old Dutch children could be explained by their word reading and listening comprehension skills two years previously after controlling for the autoregressive effect of earlier reading comprehension.

The Simple View of Reading implies that as decoding or word recognition increases, approaching the "perfect" score of 1, the importance of listening comprehension in predicting reading comprehension will increase. Results from both cross-sectional and longitudinal studies, investigating the relative contributions of decoding and listening comprehension to reading comprehension, have supported this hypothesis. In her cross-sectional study, Curtis (1980) found that the relative contribution of listening comprehension to reading comprehension increased as word identification became more automated with increasing age and word reading ability. Vellutino, Tunmer, Jaccard, and Chen (2007) also reported that listening comprehension was a more important predictor of reading comprehension for children aged 12 and 13 than for

those aged 8 and 9. Hoover and Gough's (1990) longitudinal study found decoding to be the most important predictor of reading comprehension when children were aged 7 and 8 whilst listening comprehension became the most important predictor at ages 9 and 10. Juel (1988) found that word recognition was the most important predictor for children between the ages of 7 to 10 but also found that the impact of listening comprehension increased each year. Similarly, Catts, Adlof, and Weismer (2006) found in their longitudinal study that it was not until children were aged 14 that those with impairments in listening comprehension but not decoding skills became more impaired on reading comprehension than children with poor decoding skills but adequate listening comprehension ability.

The importance of the relative contributions of decoding and listening comprehension to reading comprehension has also been shown to depend on the orthographic transparency of the language. Megherbi, Seigneure, and Ehrlich (2006) found that for French children in grades 1 and 2 (mean age 6:8 and 7:8), listening comprehension was a better predictor of reading comprehension than non-word reading ability. Similarly, de Jong and van der Leij (2002) found that, for Dutch children, their listening comprehension at age 7 was a better predictor of reading comprehension two years later than their word recognition skills at 7. Grapheme-phoneme relations are more consistent in French and Dutch than in English so word reading would be expected to become efficient enough to allow resources to be devoted to comprehension processes at an earlier age.

It should be noted that most of the studies reported here have used an additive ($D + C$) rather than a multiplicative ($D \times C$) model to account for reading comprehension in terms of listening comprehension and word recognition or decoding ability. Hoover and Gough (1990) followed 254 bilingual students from ages 7 to 10 and found that reading comprehension was best predicted by the product of nonword reading and listening comprehension at each age. Other authors, however, have not found a multiplicative model to predict reading comprehension any better than a linear, additive combination of decoding and listening comprehension (Georgiou, Das, & Hayward, 2009). Chen and Vellutino (1997) suggested a model incorporating both sum and product of decoding and listening comprehension ($R = D + C + (D \times C)$) but other authors have found that the inclusion of the product term does not enhance the

explanatory power of the model (Neuhaus, Roldan, Boulware-Gooden, & Swank, 2006; Savage, 2006; Tiu, Thompson, & Lewis, 2003). However, despite the debate about the best way to formulate the relationship between reading comprehension and its two component processes, these authors agree on the validity of the basic premise of the Simple View of Reading, the dependence of reading comprehension on listening comprehension and word recognition skills.

To summarise, the evidence reviewed here suggests that there is much empirical support for the theory that word recognition ability and listening comprehension make independent contributions to reading comprehension which change over time. In the next section evidence which potentially challenges the Simple View of Reading will be discussed.

2.5.3. Limitations

Whilst the Simple View of Reading is a parsimonious account of reading comprehension with considerable empirical support, it is nonetheless a “simple” model of an undoubtedly complex process as Hoover and Gough (1990) themselves acknowledge. It is important to be aware of the limitations in its ability to account for reading comprehension that have been exposed.

The Simple View of Reading treats decoding and listening comprehension as independent processes. Several studies, however, show considerable correlation between the two variables. Hoover and Gough (1990) found, across their whole sample, significant correlations between decoding and listening comprehension in children between the ages of 7 to 10. Correlations only became non-significant, or indeed negative, when only the poorest reading comprehenders were considered. Hagtvet (2003) found that a composite measure of word and nonword reading ability showed similar correlations with comprehension in each modality for Norwegian 9-year-olds when the comprehension task involved retelling a story. Similarly, Shankweiler et al. (1999) found significant correlations between listening comprehension and measures of decoding and word recognition ability, whilst Chen and Vellutino (1997) found that the correlation between decoding and listening comprehension became non-significant only at age 14. Furthermore, Cutting and

Scarborough (2006) showed that, although oral language skills and a composite measure of word recognition and decoding skills made independent contributions to reading comprehension, a considerable amount of variance was shared, and Vellutino, Tunmer, Jaccard, and Chen (2007) used Structural Equation Modelling to show that some of the same subskills, for example semantic skills, predicted both listening comprehension and word recognition.

Various explanations have been proposed for these findings, one of which is the “unitary phonological deficit hypothesis” (Shankweiler & Crain 1986). Poor phonological skills are known to underpin deficits in word recognition ability (Goswami & Bryant, 1990; Perfetti, Beck, Bell, & Hughes, 1987). The “unitary phonological deficit hypothesis” postulates that poor phonological skills are also related to the ability to comprehend spoken language as they reflect a deficit in the ability to retain and process verbal information in working memory. Most evidence, however, suggests that word recognition ability and reading comprehension have different determinants (Oakhill, Cain, & Bryant, 2003) and that poor reading comprehenders have normal abilities in phonological processing (Cain, Oakhill, & Bryant, 2000b; Nation, Adams, Bowyer-Crane, & Snowling, 1999; Nation, Clarke, Marshall, & Durand, 2004; Nation & Snowling, 1998b; Stothard & Hulme, 1995).

Some authors have argued that the relationship between listening comprehension and the recognition of words reflects the role of “top-down” processes in word recognition. Two routes to word reading have been proposed (Coltheart, Curtis, Atkins, & Haller, 1993). The phonological recoding route involves the use of letter-sound correspondences to convert a graphemic representation of a word into a phonological one which is used to access the meaning of the word in the mental lexicon. The lexical route allows direct access from the orthographic representation to the meaning of the word in the lexicon and is used to read familiar or irregular words. It has been argued that the reading of irregular words is supported by semantic information. Nation and Snowling (1998b) showed that children identified as poor reading comprehenders with adequate phonological skills had difficulty reading irregular words and, in another study, they demonstrated that the ability to use context to read exceptional or inconsistent words was predicted by both reading comprehension and listening comprehension after controlling for phonological skills

(Nation & Snowling, 1998a). Furthermore, Nation and Snowling (2004) found single word recognition to be predicted by semantic skills, vocabulary and listening comprehension both concurrently and longitudinally when these were individually entered into a regression analysis after controlling for phonological skills. Goff, Pratt, and Ong (2005) found irregular word reading to be a much stronger predictor of reading comprehension in children aged 8-11 than non-word reading and argued that this was because irregular word reading captures additional variance in language comprehension ability. These findings have important methodological implications as they suggest that the measure of decoding or word recognition ability used will impact on findings.

A final explanation for the relationship between decoding and listening comprehension is that it is an indirect one resulting from complex reciprocal relationships between the three skills involved. One criticism that has been levelled at the Simple View of Reading is that it implies a unidirectional relationship in which decoding and listening comprehension lead to reading comprehension (Connors, 2009). Stanovich (1986) argued that greater experience in reading itself leads to improvements in word recognition ability and in comprehension skills. In general, the children with greatest motivation to read would be the better decoders and comprehenders and these children would be expected to manifest the largest improvements in word recognition and language comprehension abilities, an example of the “Matthew effect”. Juel (1988) found that children who became good readers had a much higher exposure to print than children who became poor readers and, importantly, they made more progress in listening comprehension. Verhoeven and van Leeuwe (2008) also found relationships between reading comprehension and listening comprehension to be reciprocal especially in later grades, and concluded that the development of reading comprehension and listening comprehension is highly interdependent. Thus, the Simple View of Reading does not appear to fully account for the interdependence of word recognition, listening comprehension and reading comprehension skills.

It has also been argued that the Simple View of Reading does not capture all the complexity of the reading comprehension process since the inclusion of a third component in the model has sometimes been found to account for variance in reading

comprehension not accounted for by measures of decoding and listening comprehension. Joshi and Aaron (2000) found that a measure of processing speed, letter naming speed, explained an additional 10% of the variance in reading comprehension after controlling for the product of decoding and listening comprehension skill. They argued that this reflected the greater availability of resources for comprehension processes that resulted from the faster processing of letters and words during reading. Neuhaus, Roldan, Boulware-Gooden, and Swank (2006), however, found that letter naming speed did not predict extra variance in reading comprehension having accounted for decoding ability and listening comprehension, a finding replicated in Lee and Wheldall's (2009) study of young Malaysian readers. Johnston and Kirby (2006) did replicate Joshi and Aaron's findings but only when the measure of decoding was a non-word reading task rather than a word recognition task. They hypothesised that naming speed contributes to reading comprehension through its relationship with lexical access in word recognition. Therefore, its effect is already included in the product of word recognition and listening comprehension. Furthermore Tiu, Thompson, and Lewis (2003) found that variance in reading comprehension explained by processing speed was itself explained by IQ and argued that IQ accounts for variance in reading comprehension over and above that explained by decoding and listening comprehension. Connors (2009), on the other hand, found that neither IQ nor a measure of processing speed, articulation speed, explained additional variance in reading comprehension after statistically controlling for word recognition and listening comprehension, although additional variance was explained by attentional control. Connors argued that reading comprehension requires the co-ordination of word recognition and language comprehension processes.

The role of vocabulary as a possible third component has also been investigated. Ouellette and Beers (2010) found that a measure of vocabulary depth, the ability to provide oral definitions of words, made a unique contribution to reading comprehension in 12-year-olds, but not in 7-year-olds, having controlled for listening comprehension, phonological awareness, decoding and irregular word recognition. Other research, however, used confirmatory factor analysis to show that in adults with poor literacy skills, reading comprehension was not significantly related to vocabulary when the factors of word recognition and listening comprehension were also in the

model (Sabatini, Sawaki, Shore, & Scarborough, 2010). These authors concluded that word recognition and listening comprehension were primary factors in reading comprehension, as suggested by the Simple View of Reading.

This review of the literature examining possible third components suggests that very little has yet been established. Nevertheless, the research surveyed does give an indication that the Simple View of Reading might not provide a complete account of the components of reading comprehension.

Whilst the limitations discussed in this section so far suggest that the Simple View of Reading might not capture all the complexity of reading comprehension, more profound concerns have been raised regarding the assumption that the processes involved in reading comprehension are essentially the same as those involved in listening comprehension, once word recognition has been accounted for. Obviously, the structure of normal spoken language differs to that of written language (Chafe & Danielewicz, 1987). Similarly, the contexts in which listening comprehension occurs are often interactive, providing opportunities for listeners to clarify their interpretation of the speaker's meaning. No such opportunities are available when reading comprehension is taking place (Rubin, 1980). However, Hoover and Gough (1990) made it clear that the Simple View of Reading could only be properly assessed using comparable materials to measure comprehension of spoken and written texts, and the studies reviewed above have adopted this approach.

Nevertheless, it has been argued that simply changing the modality of presentation of a text changes the processing demands of the comprehension task (Rubin, 1980). In listening comprehension, the use of prosodic cues facilitates access to lexical, syntactic and discourse information (Rubin, 1980). For example, pauses occurring at syntactic boundaries increase the comprehensibility of the text (Sticht, 1972). Use of stress helps organise the discourse by, for example, disambiguating a pronoun's referent. Intonation is also useful, for example in clarifying when a question is being asked. Children have been shown to be more dependent on prosody for syntactic processing than adults (Schreiber, 1987). Although written text does contain punctuation, children have to learn how to use this. Nevertheless, the permanence of written text can be seen as compensating for the lack of prosodic information (Rubin,

1980). In normal reading situations, readers can reread portions of the text to aid comprehension and they can also look ahead in the text. Listeners on the other hand cannot, in normal listening situations, re-visit the text. Listeners and readers, therefore, make decisions about meaning based on different information (Danks & End, 1987). Furthermore, the rate of listening is controlled entirely by the speaker whilst the rate of reading is under the comprehender's control (Danks & End, 1987). In listening, comprehenders have to process the stimuli immediately whether or not they are still processing the preceding material.

Critics of a “unitary view” of comprehension processes acknowledge that listening comprehension and reading comprehension are similar in that they both involve language processes. However, they argue that the different cognitive demands made by reading comprehension and listening comprehension mean that different processing strategies may be employed in the two modalities (Danks & End, 1987). Rubin (1980), for example, argues that children need to learn how to use the permanence of text strategically in order to compensate for a lack of prosody. Similarly, Kirby and Savage (2008) speculate that there is more opportunity for the use of meta-cognitive strategies, such as locating information and finding main ideas, in reading comprehension than in listening comprehension because, when comprehending spoken text, the use of such strategies would potentially interfere with the reception of information.

There is some experimental evidence to support the hypothesis that different strategies are used in listening comprehension and reading comprehension. Hildyard and Olson (1978) presented passages to children aged 9 or 12 in either written or auditory modalities. Children were then tested on their recognition of literal text ideas and on their recognition of inferences necessary for text coherence. It was found that, after listening to stories, children's recognition for inferences necessary for text coherence was better than their recognition of explicitly stated text ideas. However, after reading stories, the reverse was true. The authors concluded that children engage in different strategies when listening and reading, paying close attention to the exact content of written text but focussing on the formation of a coherent representation of the gist of spoken discourse because it cannot be re-visited.

An implication of these ideas is that, given adequate word recognition ability, the comprehension of written text may be higher than that of its spoken equivalent. This phenomenon is theoretically impossible according to the Simple View of Reading which proposes that listening comprehension effectively sets a limit on reading comprehension. Durrell (1969) presented the same material in oral and written form to children aged between 7 and 14. Whilst listening comprehension was more effective for the younger children, in the oldest group reading comprehension was superior to listening comprehension. Similarly, Horowitz and Samuels (1985) found that, when reading expository material that was well within their reading capabilities, good readers aged 12 showed higher comprehension scores than when they listened to the material.

Several studies have compared performance on tests of reading comprehension and listening comprehension using the Sentence Verification Technique, developed by Royer and colleagues. In this test, children listen to or read a passage followed by test sentences of 4 types – originals and paraphrases, meaning changes and distractors. Children have to correctly identify sentences which express ideas from the passage which they have read or to which they have listened. Studies using this technique have shown that good readers perform better when material is read rather than heard as long as the material is easy for them to read (Royer, Kulhavy, Lee, & Peterson, 1986; Royer, Sinatra, & Schumer, 1990). The authors suggested that readers were able to take advantage of the opportunity to proceed at their own pace and read material in the order of their own choosing when texts were easy to read. However, for the difficult texts, listening overcame word reading problems and prosodic cues helped with syntactical analysis of the more complex material.

Carlisle and Felbinger (1991) also used the Sentence Verification Technique in their cross-sectional study of children aged 10, 12 and 14. Four types of comprehender were identified, those who were generally good, those who were generally poor and those who had a specific deficit in reading comprehension or in listening comprehension. The identification of children who performed poorly on the listening comprehension test, despite demonstrating adequate reading comprehension ability, led Carlisle and Felbinger to argue against the use of listening comprehension as a measure of potential in reading comprehension. Furthermore, Carlisle and Felbinger

examined patterns of performance of all the comprehension groups on the four different types of test sentence. They ascertained that the children who had a specific problem in listening comprehension tended to adopt a strategy of paying insufficient attention to exact wording and of processing sentences for general ideas. Interestingly, for all groups, the pattern of errors on the four different types of sentence was different in the two modalities. This led Carlisle and Felbinger to conclude that the memory representations of text ideas and the processes that the children were using to understand text seemed to differ according to the modality of presentation of information.

The relationship between reading comprehension and listening comprehension has also been shown to depend on the type of text involved, a factor which is not taken into account in the Simple View of Reading. Using the Sentence Verification Technique, Diakidoy, Stylianou, Karefillidou, and Papageorgiou (2005) replicated previous findings for narrative texts, demonstrating that listening comprehension was more efficient than reading comprehension in the early grades, but that the reverse was true at age 14. However, the same pattern was not obtained for expository texts. Listening comprehension of expository text was not more efficient than reading comprehension at any age, and listening comprehension did not predict reading comprehension for expository material. Children learning to read will have had less experience of oral expository text than of oral narrative text. Also, for the comprehension of expository texts, the advantages conferred by written presentation may be most useful. For a complicated passage, including unfamiliar material, good comprehension may depend on having the opportunity to control the rate of processing of information as well as on the ability to re-process sections for coherence purposes.

To summarise this section, the Simple View of Reading has been shown to be unable to explain some empirical findings from studies investigating the relationship between reading comprehension and listening comprehension. In the next section, the implications of the Simple View of Reading for research and policy are outlined.

2.5.4. Implications

The Simple View of Reading has important implications for classroom practitioners. The “Rose Report” into the teaching of early reading advocated the adoption of the Simple View of Reading as a conceptual framework on which to base the teaching of reading in schools (Stuart & Stainthorp, 2006). This recommendation was enshrined in the Primary Framework for Literacy (DfES, 2006) and, since 2007, practitioners have been advised to teach word recognition and listening comprehension skills in order to ensure reading comprehension success.

Furthermore, the Simple View of Reading has been used to challenge the use of tests of IQ in the identification of dyslexia (Savage, 2001; Spring & French, 1990). It has been argued that listening comprehension ability is a better indication of potential reading comprehension ability than the traditional psychometric tests used and that specific decoding problems can be identified by comparing comprehension in the two modalities.

The Simple View of Reading has also been useful in directing research activity by identifying comprehension processes in reading as important. Although the purpose of reading is comprehension of the material read, most research in reading has been concerned with word recognition ability only (e.g. Goswami & Bryant, 1990). The Simple View of Reading has placed comprehension processes at the heart of a conceptualisation of reading and offers an explanation as to how the relative contributions of word reading and comprehension processes change in the course of reading development. As such, the Simple View of Reading has provided a useful framework for researchers investigating the nature of reading comprehension and its development.

2.5.5. Rationale for current research

The evidence reviewed in the preceding sections suggests that there is considerable empirical support for the Simple View of Reading and that it has important implications for policy and practice as well as for research. However, concerns about

the Simple View of Reading have also been raised, some of which will be addressed in the work presented in this thesis.

Of particular interest is the issue of whether or not there are differences in the comprehension processes used when reading and listening due to the different demands made by comprehension in the two modalities. Although some studies have suggested that the processes used in the two modalities are not exactly the same (Hildyard & Olson, 1978; Carlisle & Felbinger, 1991), this area has received little research attention. Furthermore, whilst reading comprehension has been relatively well researched, the same cannot be said for listening comprehension. Reading comprehension research has emerged as part of a larger body of research concerned with reading and, as will be illustrated in the next chapter, much is now known about the component skills involved at word-, sentence- and discourse-level. Much is also known about the role of working memory and the importance of background knowledge. In contrast, very little is known about the component skills involved in listening comprehension. This may be due, in part, to the assumption that listening comprehension and reading comprehension involve the same processes and will be underpinned by the same skills.

It seems strange that little research has focussed specifically on the ability to understand spoken discourse. From a theoretical point of view, it has been acknowledged that there are differences in task demands in comprehension in the two modalities (Stuart, Stainthorp, & Snowling, 2008). Gough and Tunmer (1986) asserted that exactly the same processes are involved in reading comprehension and listening comprehension. However, Hoover and Gough (1990) subsequently stated that the Simple View of Reading does not claim that the processes employed in comprehension in the two modalities are exactly the same. Instead, these authors suggested that any differences between comprehension processes in the two modalities are minor compared with the similarities between them.

The issue of whether or not reading comprehension and listening comprehension involve the same processes has practical as well as theoretical significance. Estimates suggest that children spend a considerable proportion of their school day engaged in listening activities (Lehto & Anttila, 2003). Over the past three decades, researchers

have been calling for specific instruction in listening comprehension in school. Curtis (1980) regarded listening comprehension as important not only for the development of reading comprehension skills but also for the opportunity it provides for children with poor word recognition skills to engage in comprehension. Similarly, Juel (1988) argued that poor readers need much practice in listening comprehension so that they do not fall behind in the acquisition of vocabulary and concepts. Furthermore, the introduction of the Simple View of Reading into the school curriculum has highlighted the role that listening comprehension ability plays in the development of early reading comprehension skills (Stuart & Stainthorp, 2006). Yet there is little research evidence, specific to listening comprehension, on which to base listening comprehension instruction.

To summarise, there seem to be two possible ways to describe the relationship between reading comprehension and listening comprehension. It is possible that, as suggested by Gough and Tunmer (1986), reading comprehension involves exactly the same processes as listening comprehension once words have been recognised. If this is the case, the investigation of comprehension in one modality will yield results which can be “extrapolated” to comprehension in the other modality. Alternatively the different demands of comprehension in the two modalities may mean that different cognitive-linguistic processes are important in each (Carlisle & Felbinger, 1991; Danks & End, 1987; Hildyard & Olson, 1978). If this view is correct, it will be necessary to establish separate evidence bases for comprehension in each modality. The empirical work presented in this thesis aimed to distinguish between these two possibilities.

In the following chapter, existing research evidence is reviewed concerning the component skills of comprehension.

Chapter 3.

Components of comprehension

3.1. Introduction

As explained in Section 2.3, the successful construction of a situation model of a text involves processing at different levels. This chapter reviews the research evidence for the involvement of word-, sentence- and discourse-level skills and processes in comprehension. Specifically, at the word-level, it considers the role played by word recognition skill and by vocabulary and semantic skills. At the sentence-level, the importance of syntactic skills is discussed, whilst, at the discourse-level, evidence for the involvement of inferencing in comprehension is reviewed. This chapter also considers the role played in comprehension by the amount of exposure to print and by the cognitive abilities of working memory and general intelligence. For reasons given in the previous chapter, most research has considered component skills of reading comprehension, rather than listening comprehension, and this is reflected in the fact that most of the research findings discussed here have emerged from studies of reading comprehension. Where listening comprehension has been directly investigated, research evidence is provided.

Most of the research reviewed here has taken place within an individual differences framework. Some studies have used correlational and regression techniques to explore the predictors of comprehension in relatively unselected samples of children (e.g. Cain, Oakhill, & Bryant, 2004; Goff, Pratt, & Ong, 2005). Other studies have used a quasi-experimental approach in which groups of “good comprehenders” and “poor comprehenders” are compared (e.g. Cain & Oakhill, 1999; Oakhill, 1984). Typically, a poor comprehender has age-appropriate word recognition skills, but a reading comprehension age well below both their chronological and reading accuracy ages. A good comprehender, on the other hand, has age-appropriate word reading skill and a reading comprehension age at or above their reading accuracy age. Thus, the groups of good and poor comprehenders are selected so that they are matched for word recognition ability but differ on reading comprehension. By identifying the skills and knowledge on which children with poor comprehension differ from children with

good comprehension, the factors important in reading comprehension can be identified. In line with the convention used in the literature, this review uses the term “poor comprehender” to refer to a child with a reading comprehension deficit in the absence of problems with decoding or word recognition ability.

3.2. Word-level skills

3.2.1. Word-recognition/ decoding ability

The ability to identify words is obviously a prerequisite for reading comprehension. Yuill and Oakhill (1991) report that, in the literature, correlations ranging from 0.6 to 0.8 have been found between reading comprehension and word recognition ability. Interestingly, however, much research on reading comprehension excludes children with poor word recognition skills, focussing instead on “poor comprehenders”, children who fail to comprehend what they read despite having adequate word recognition skills.

Nevertheless, it has been argued that reading comprehension deficits are due primarily to deficits in word recognition ability (Perfetti, 1985). According to his “verbal efficiency theory”, word recognition and reading comprehension processes compete for limited processing resources. Processing capacity necessary for text comprehension is used up by effortful word reading. Perfetti, Marron, and Foltz (1996) argue that even when good and poor comprehenders are matched on the accuracy of their word recognition ability, the word processing skills of the poor comprehenders may be slow and inefficient and this may affect their comprehension ability. However, Stothard and Hulme (1996) failed to find any difference in text reading speed of 7-8 year-old good and poor comprehenders matched for word recognition ability. Similarly, Goff, Pratt, and Ong (2005) did not find text reading speed to explain any variance in reading comprehension ability in a sample of 8-10 year-olds, once the contribution of irregular word reading had been controlled for.

As mentioned in the previous chapter, some research has found word recognition ability to be correlated with listening comprehension (Hagtvet, 2003; Shankweiler et al., 1999). A full discussion of the possible reasons for this has been given in Section 2.5.3.

3.2.2. Vocabulary and semantic knowledge

If language is to be understood, word meanings need to be accessed. High correlations between receptive vocabulary and reading comprehension have been reported for both adults (Braze, Tabor, Shankweiler, & Mencl, 2007) and children (Biemiller, 2003). Instruction in vocabulary has been shown to lead to improvements in reading comprehension in children (Beck, Perfetti, & McKeown, 1982).

Longitudinal studies have suggested a predictive role of vocabulary for later reading comprehension. Muter, Hulme, Snowling, and Stevenson (2004) found a moderate correlation between receptive vocabulary in kindergarten and reading comprehension 2 years later whilst Senechal, Ouellette, and Rodney (2006) showed that receptive vocabulary measured in kindergarten predicted unique variance in reading comprehension 3 years later after controlling for early literacy skills and parental education and literacy levels. Furthermore, vocabulary has been shown to explain variance in reading comprehension after accounting for the autoregressive effect of earlier reading comprehension (de Jong & van der Leij, 2002; Seigneuric & Ehrlich, 2005).

Within the literature, a distinction is made between vocabulary *breadth*, a measure of the number of items held in the mental lexicon that have some meaning to a child, and vocabulary *depth*, a measure of the richness of knowledge about these words (Tannenbaum, Torgeson, & Wagner, 2006). It has been argued that it is not vocabulary breadth that determines reading comprehension skill, but vocabulary depth as this influences speed and efficiency of semantic access (Nation & Snowling, 1999; Ouellette, 2006). These authors argue that, as with word recognition ability, if semantic access is slow and effortful, fewer cognitive resources will be available for comprehension. Efficient semantic access depends on organisation within the semantic system. Within the mental lexicon, phonological representations of words are stored with connections to semantic representations (Ouellette, 2006). Efficient semantic processing occurs when a rich semantic network exists and an interconnected knowledge base can be accessed (Beck, Perfetti, & McKeown, 1982).

A commonly used measure of vocabulary depth, or semantic knowledge, is the ability to orally define words. Children who are poor reading comprehenders have been shown to perform less well on this task than good comprehenders (Nation, Clarke, Marshall, & Durand, 2004). The ability to orally define words has also been shown to explain variance in reading comprehension having accounted for age, non-verbal intelligence, phonological skills, and regular and exception word reading (Ricketts, Nation, & Bishop, 2007). Furthermore, Ouellette (2006) reported that the ability to provide word definitions explained additional variance in reading comprehension after statistically controlling for receptive vocabulary breadth, whilst receptive vocabulary did not explain additional variance in reading comprehension having controlled for the ability to define words. It should be noted, however, that findings in this area are inconsistent. Using structural equation modelling, Tannenbaum et al. (2006) found vocabulary breadth to be a better predictor of reading comprehension in 9-year-old children than vocabulary depth.

Nation, Snowling and colleagues have conducted a series of experiments demonstrating the relationship between children's reading comprehension skills and various other measures of semantic knowledge. They found that poor reading comprehenders were less accurate and slower on a synonym judgement task than good comprehenders, and also performed worse on a word association task measuring semantic fluency (Nation & Snowling, 1998b). These authors also used an auditory lexical decision task to show that, compared to good comprehenders, poor comprehenders were sensitive to semantic relations between functionally related pairs of words, but insensitive to more abstract semantic relations based on category membership (Nation & Snowling, 1999). Nation and Snowling (2004) used regression techniques to show that reading comprehension was concurrently predicted by semantic skills, measured by word association and synonym judgement tasks, after statistically controlling for age, non-verbal ability and phonological skills. Furthermore, these semantic skills were found to predict reading comprehension 4.5 years later, even after taking account of the autoregressive effects of earlier reading comprehension.

Cain, Oakhill, and Lemmon (2005), however, found that that 9-year-old good and poor comprehenders, matched on word reading accuracy and sight vocabulary,

performed equally well on a word association test of semantic fluency. Also, poor comprehenders did not differ from good comprehenders on their ability to provide appropriate meanings for novel transparent idioms, for example “to run around like scalded pigs”. This finding was interpreted as indicating that poor comprehenders’ semantic analysis skills were not impaired. This raises an interesting methodological issue. Nation and colleagues used non-word reading accuracy as a measure of decoding skill when selecting groups of good and poor comprehenders, whilst Cain et al. (2004) selected children on the basis of their adequate word recognition skills. As demonstrated in the previous chapter, word recognition is itself affected by “top-down” semantic processes. Therefore, a selection procedure based on word reading accuracy may effectively “screen out” children with poor semantic knowledge (Cain, 2006).

Longitudinal studies have demonstrated that the relationship between vocabulary and reading comprehension is reciprocal (Bast & Reitsma, 1998; Verhoeven & van Leeuwe, 2008). Good vocabulary knowledge enables text to be understood but the experience of understanding text itself leads to vocabulary development. In fact it has been suggested that vocabulary acquisition and reading comprehension may be underpinned by shared processes. Sternberg and Powell (1983) argued that the ability to work out the meanings of unfamiliar vocabulary items depended on inferencing skill, also important in reading comprehension as will be shown later. Cain, Oakhill, and Lemmon (2004) provided empirical support for this hypothesis by demonstrating that poor reading comprehenders were poor at using contextual cues to infer new word meanings compared to good reading comprehenders.

Compared to the number of studies which have considered the role of vocabulary in reading comprehension, very few have investigated its relationship with listening comprehension. In pre-school children, a correlation between receptive vocabulary and listening comprehension has been reported (Adams, Bourke, & Willis, 1999; Florit, Roch, Altoe, & Levorato, 2008).

Some studies have directly compared the contribution made by vocabulary to reading comprehension and listening comprehension. For example, Burgoyne, Kelly, Whiteley, and Spooner (2009) used regression methods to compare the ability of

expressive and receptive vocabulary to predict listening comprehension and reading comprehension in both monolingual children and children with English as an Additional Language (EAL). For the monolingual children, expressive vocabulary uniquely predicted listening comprehension whilst receptive vocabulary uniquely predicted reading comprehension. The authors argued that this reflected the different demands made by comprehension in the two modalities. In the reading comprehension test, children could re-visit the text and use the vocabulary and information it contained to help them answer questions. In the listening comprehension test, where this was not possible, the demands on expressive vocabulary were much higher. Interestingly, expressive vocabulary but not receptive vocabulary was a unique predictor of both reading comprehension and listening comprehension for children learning EAL suggesting that comprehension processes are not the same for both groups of children.

Other studies comparing reading comprehension and listening comprehension have found that vocabulary and semantic knowledge have greater importance in reading comprehension than in listening comprehension and it has been suggested that this could reflect the influence of semantic knowledge on word reading itself. This point was made by Braze et al. (2007) who found that a composite measure of receptive and expressive vocabulary explained additional variance in the comprehension of written sentences after statistically controlling for the comprehension of spoken sentences. The composite measure did not, however, explain additional variance in the comprehension of spoken sentences after statistically controlling for the comprehension of written sentences. Similarly, Hagtvet (2003) compared reading comprehension and listening comprehension in 9-year-olds when the task involved the retelling of a story. After taking into account syntax and phonemic awareness, the unique variance in listening comprehension explained by the ability to orally define words was just short of statistical significance. However, the ability to orally define words was a unique predictor of reading comprehension ability assessed in the same way. These are interesting findings which suggest that vocabulary and semantic knowledge may not operate in exactly the same way in reading comprehension and listening comprehension.

3.3. Syntactic skills

Studies that have explored the role of syntactic skills in reading comprehension fall into two categories (Cain & Oakhill, 2007). Some have investigated syntactic *knowledge*, the implicit knowledge that enables meaning to be constructed from sentences (e.g. Stothard & Hulme, 1992). Others have investigated syntactic *awareness*, an explicit understanding of the rules of grammar that facilitates conscious reflection on grammatical structures (e.g. Nation & Snowling, 2000). Both types of study will be reviewed here.

There is some evidence that syntactic skills in children impact on reading comprehension indirectly through their relationship with word recognition ability. Several studies have demonstrated a correlation between word recognition skill and a measure of syntactic awareness (Bowey, 1986) or of syntactic knowledge (Hagtvet, 2003). Tunmer and Hoover (1992) argued that understanding of syntactical constraints aids in the reading of unfamiliar words in sentences and reported that syntactic awareness explained additional variance in word recognition ability after statistically controlling for phonological awareness.

There is, however, evidence to suggest that syntactic skills are also directly related to reading comprehension. Having recognised the words and retrieved their meanings, it would seem that syntactic knowledge is necessary to establish the correct representation of a sentence. Stothard and Hulme (1992) found that poor comprehenders performed worse than good comprehenders, matched on age, word recognition ability and receptive vocabulary, on a test of receptive syntactical skills in which children had to identify the picture that matched a sentence they had heard. Similarly, using regression techniques, Goff, Pratt, and Ong (2005) found receptive grammar skills explained additional variance in reading comprehension in children aged 9 to 11 after accounting for irregular word reading and receptive vocabulary.

A relationship between reading comprehension and measures of syntactic awareness has also been reported. Bowey (1986) found that the ability to correct grammatically deviant sentences continued to predict reading comprehension after controlling for word recognition skill and vocabulary. She also found that the ability to correct grammatically deviant sentences correlated with measures of on-going comprehension

monitoring, leading her to suggest that syntactic awareness influences reading comprehension through facilitating identification and correction of errors. Nation and Snowling (2000) found that good and poor reading comprehenders, matched for age, non-word reading and non-verbal IQ, differed in their ability to re-order aurally presented jumbled sentences. The poor comprehenders were particularly impaired when re-ordering passive, rather than active, sentences and sentences that were semantically ambiguous. Poor comprehenders have also been shown to be worse than good comprehenders at repeating sentences of increasing grammatical complexity and at replacing verbs in the present tense with their past tense equivalents (Nation, Clarke, Marshall, & Durand, 2004).

The role of syntactic awareness in reading comprehension has also been explored in a longitudinal study investigating the development of French children from ages 5-8 (Demont & Gombert, 1996). It was found that syntactic awareness predicted later reading comprehension after controlling for intelligence and vocabulary.

Some authors have suggested that the relationships between syntactic skills and both word recognition and reading comprehension abilities reflect the dependence of all these variables on phonological skills. Gottardo, Stanovich, and Siegel (1996) found that, in 9-year-old children, syntactic awareness, measured by the ability to identify and correct syntactic errors in sentences, failed to predict reading comprehension or word reading ability once verbal WM and phonological awareness had been accounted for. The authors claimed that poor syntactic skills are often associated with reading comprehension problems because impairments in both are due to deficiencies in phonological processing skills.

It would seem that, to date, only one study has been published in which the role of syntactic skills in the comprehension of extended spoken texts has been investigated. Hagtvet (2003) measured the syntactic knowledge of Norwegian 9-year-olds using a test in which the children had to identify the picture which matched a spoken sentence. After controlling for phonemic awareness and vocabulary, the contribution of syntactic skills approached significance ($p = .06$) when the comprehension task involved retelling a story that had been heard. When the same comprehension task was presented in written form, however, syntactic skills were not significant. This

difference possibly reflects the greater contribution of phonemic awareness to reading comprehension and illustrates why it is important to assess the contribution of component processes of comprehension when the measure of comprehension used is not confounded with word reading ability.

3.4. Inferencing

As discussed in Sections 2.2 and 2.3, effective comprehension requires the construction of a situation model that goes beyond what is explicitly stated in a text. According to the Construction-Integration model, inferencing is a critical component of comprehension (Kintsch, 1998; Kintsch & Rawson, 2005). There is much empirical evidence to show that inferencing is involved in reading comprehension in adults (e.g. Hannon & Daneman, 1998; Long, Golding, & Graesser, 1992; Potts, Keening, & Golding, 1988).

McKoon and Ratcliff (1992) define an inference as “any piece of information that is not explicitly stated in a text” (p.440). This definition incorporates a huge range of inferences. Graesser, Singer, and Trabasso (1994) have identified 13 classes of inferences, ranging from those that are relatively simple to those that are complex and elaborative. For example, according to Graesser et al., a simple referential inference is made when the relationship between a pronoun and its referent is encoded, and is necessary when comprehending a statement such as “...on removing the *fork* the eye came with *it*”. The inference that “*it*” refers to “*the fork*” is necessary to link different parts of the text and maintain coherence. A more complex, elaborative inference is made, however, on reading “...he stuck a pickle-fork into his right eye, and on removing the fork the eye came with it”, when the comprehender infers that this action resulted in the character becoming blind in his right eye. To make this inference, background knowledge is required regarding the role of the eye and the consequences of its removal. Graesser et al. refer to this inference as a “causal consequence”. Other inferences include those regarding spatial setting and layout, character traits and emotional states.

It is important to distinguish between those inferences which can be drawn from information given in the text and those which require the comprehender to access

information or knowledge which is not stated explicitly in the text. Inferences are termed “explicit” or “text-connecting” when all the information that is necessary for inferencing is provided in the text. Explicit inferences are deliberate deductions (Oakhill, 1982). Inferences that require the integration of material in the text with the comprehender’s background knowledge are referred to as “knowledge-based” or “implicit” inferences. Broadly speaking, inferences can be categorised into those that are necessary to establish coherence and those that provide additional information not strictly necessary for text coherence. Referential inferences fall into the first category, as do other inferences referred to as “bridging”, “backward” or “gap-filling”. Inferences belonging to the second category are described as “predictive”, “forward” or “elaborative”. It is not always clear, however, which category a particular inference belongs to (Long, Golding, & Graesser, 1992).

There is an enormous amount of debate as to the inevitability of “implicit”, knowledge-based inferences. Proponents of the “minimalist” or “memory-based” view argue that the only inferences generated “on-line” during comprehension are those required for local coherence, for example referential inferences, or based on easily available information. All other inferences are generated “off-line” when they are necessary to perform a task, for example in response to a question, and result from strategic, problem-solving processes that operate on the textbase that has been formed (McKoon & Ratcliff, 1992). Proponents of the “constructionist” position, however, argue that many more elaborative inferences are generated during skilled reading than are necessary for local text coherence. These include causal links between events and actions, and the goals and motivations of characters (Graesser, Singer, & Trabasso, 1994). On-line priming tasks have been used to show that elaborative goal-related inferences are generated whilst reading (Long, Golding, & Graesser, 1992). They have also shown that reading comprehension in adults is related to the ability to generate knowledge-based inferences on-line but is not related to the ability to form an accurate textbase representation, a representation of the literal content of the text (Long, Oppy, & Seely, 1994).

There is a large amount of evidence, provided by Oakhill and her colleagues, to suggest that the inability to generate inferences is associated with poor reading comprehension in children. Furthermore, this evidence suggests that problems with

inferencing can be evident in children who are able to form an adequate representation of the literal content of the text. The studies conducted by Oakhill and her colleagues have investigated the role of both explicit and implicit inferences in children's reading comprehension and are reviewed here.

Oakhill (1982) investigated children's ability to integrate information across sentences, i.e. to make explicit text-connecting inferences. Children listened to short stories consisting of three sentences. For example, they heard "The car crashed into the bus. The bus was near the crossroads. The car skidded on the ice." Children were then given a recognition test in which they were shown two original sentences, one semantically congruent foil (e.g. "The car was near the crossroads") and one semantically incongruent foil (e.g. "The bus skidded on the ice"). There was no difference in the ability of good and poor comprehenders, matched on word reading skill and sight vocabulary, to correctly identify the original sentences. Good reading comprehenders, however, were more likely than poor comprehenders to incorrectly "recognise" the semantically congruent foil. Oakhill claimed that only good comprehenders had formed a situation model of the text by integrating information across sentences, despite the fact that both groups appeared to have formed an accurate textbase representation of the discourse. These findings, however, have not been replicated by Spooner, Gathercole, and Baddeley (2006) who found no difference in the integration abilities of skilled and less skilled comprehenders.

Another strand of Oakhill's research has focussed on children's ability to make referential inferences, specifically by processing anaphors, words which take their meaning from an earlier part of the text. For example, in one study, children were presented with sentences made up of two clauses, such as "Peter lent ten pence to Liz because she was very poor" (Oakhill & Yuill, 1986). Children were then asked a question such as "Who was very poor?". To answer this question correctly, children had to infer that "she" referred to "Liz" as Liz was the only female antecedent in the main clause. However, children were also given the task when no gender cue was present. For example, they were asked "Who was very poor?" after hearing "Peter lent ten pence to John because he was very poor". With no gender cue present, children had to use their background knowledge of the situation in which the lending of money might occur in order to infer that John was the poorer of the two characters.

Oakhill and Yuill found that less skilled comprehenders performed more poorly on the task than skilled comprehenders whether the gender cue was present or not. Furthermore, in a second experiment, Oakhill and Yuill demonstrated that less skilled comprehenders were poorer than skilled comprehenders when completing a sentence by inserting a correct anaphor (either “he” or “she”). The difference between the performance of good and poor comprehenders was particularly great when sentence completion required a complex inference to be made. For example, to complete the sentence “Steven gave his umbrella to Penny in the park because.....wanted to keep dry” children had to use their knowledge that umbrellas keep people dry to infer that Penny was given the umbrella because she wanted to stay dry. Based on their findings, Oakhill and Yuill concluded that poor comprehenders’ difficulty with anaphoric resolution has two possible sources. The first is that poor comprehenders do not attend properly to pronouns and do not use syntactic cues, such as gender, effectively. The second is that, in the absence of syntactic cues, poor comprehenders do not make the knowledge-based inferences required to relate information from different parts of a text.

The broader role of the ability to make implicit knowledge-based inferences in comprehension has been investigated extensively by Oakhill and her colleagues. Oakhill (1984) asked good and poor comprehenders to read four short stories. After each story, the children were asked questions, some of which required them to recall information explicitly stated in the text whilst others required them to make a knowledge-based inference. For example, in one of the stories, children read “He picked up his two books and put them in a bag. He started pedalling to school as fast as he could.” At the end of the story they were asked “How many books did John pick up?”, a question which simply requires children to remember information from the text. They were also asked “How did John travel to school?” To answer this question correctly, children had to integrate material they had read in the text, i.e. that John travelled to school by pedalling, with their knowledge that children pedal when they are on bikes. Oakhill found that when answering the questions from memory, poor comprehenders performed worse than good comprehenders on both sorts of questions. However, when children were given the text to refer to, poor comprehenders performed worse than good comprehenders on questions requiring inference generation only. Oakhill concluded that poor comprehenders had problems making

knowledge-based inferences and that this difficulty did not arise simply because they had a poorer memory for the text. Further research conducted by Cain and Oakhill (1999) investigated both text-connecting and knowledge-based inference generation and found that, after reading a short story, poor comprehenders were worse than good comprehenders at answering questions requiring either type of inference when the story could not be referred to. There was, however, no difference between the groups in the ability to answer questions requiring the recall of factual information from the stories, suggesting that both groups were able to form an adequate textbase representation. When the story was present, poor comprehenders performed worse than good comprehenders on questions requiring knowledge-based inferences only. Results from regression analyses also suggest comprehension ability is determined by inferencing ability rather than by the ability to form an adequate textbase representation. Oakhill, Cain, and Bryant (2003) found that, having statistically controlled for age, non-verbal IQ, verbal IQ, and receptive vocabulary, the ability to answer questions requiring the generation of text-connecting or knowledge-based inferences continued to explain unique variance in 8-9 year-olds' reading comprehension ability, but literal memory for the text did not.

In order to make a knowledge-based inference, it is obviously necessary for a comprehender to possess the relevant knowledge. It is possible, therefore, that children who fail to make knowledge-based inferences simply lack the appropriate knowledge. In their study, Cain and Oakhill (1999) investigated this possibility and found that children failing to make knowledge-based inferences did possess the general knowledge needed to do so but failed to make inferences spontaneously. The authors hypothesised that whilst good comprehenders strive for coherence when reading, poor comprehenders may simply be aiming to read accurately. In a later study of inferencing, Cain, Oakhill, Barnes, and Bryant (2001) attempted to control for individual differences in general knowledge by using a paradigm developed by Barnes, Dennis, and Haefele-Kalvaitis (1996). Children were taught a novel knowledge base, a series of facts about an imaginary planet "Gan", and then listened to a multi-episode story concerning events on "Gan". After each episode they were asked questions regarding literal and inferential material from the episode. The poor comprehenders were worse at answering literal questions than the good comprehenders. Inferential questions involved the retrieval of information from the

novel knowledge base and its integration with material in the text. Recall of the knowledge base was assessed at the end of the story and answers to questions involving inferences were only included in the analysis if the relevant part of the knowledge base was recalled. When knowledge was controlled for in this way, poor reading comprehenders were found to generate fewer inferences than good comprehenders. The authors concluded that lack of knowledge was not the primary source of poor comprehenders' inferencing problems.

Another possible explanation for the relationship of inferencing ability with reading comprehension is that it is mediated by working memory (WM). As will be discussed in Section 3.6, WM is widely believed to be implicated in reading comprehension problems. It could be argued that inferencing itself requires recently read propositions to be held in WM and integrated with previously processed material or material retrieved from long-term memory. Cain, Oakhill, and Bryant (2004) investigated this possibility and found that the ability to make text-connecting and knowledge-based inferences predicted variance in reading comprehension after controlling statistically for word reading accuracy, sight vocabulary, receptive vocabulary and verbal WM span. Thus, it would appear that inferencing ability is not wholly mediated by WM.

The results described above suggest a relationship between reading comprehension and inferencing ability, but do not specify a direction of causality. However, to investigate this issue, Cain and Oakhill (1999) included in their study a third group of younger children, matched with the poor comprehenders on comprehension ability. These children performed better than the poor comprehenders on questions requiring text-connecting inferences, indicating that, as far as this type of inference was concerned, inferencing ability was not a by-product of comprehension skill itself. Other evidence suggesting that problems with inferencing may be a cause rather than a consequence of reading comprehension impairments comes from training studies. These show that poor comprehenders benefit more than good comprehenders from being taught how to make inferences from "clue" words and how to generate questions to test their understanding (McGee & Johnson, 2003; Yuill & Joscelyne, 1988; Yuill & Oakhill, 1988).

The research discussed here suggests, therefore, that the ability to generate inferences is an important predictor of reading comprehension success in children which cannot be explained entirely by WM, lack of knowledge or the ability to form a textbase representation of the discourse. It would seem that differences in inferencing ability arise from differences in aims and strategies. Good comprehenders have higher “standards of coherence” than poor comprehenders and are more aware of when and how it is appropriate to use general knowledge to make texts comprehensible (Cain & Oakhill, 1999). An interesting finding is that although there is a consensus that having access to an accurate textbase is not sufficient for inferencing to occur, the studies described above vary in whether they find a relationship between reading comprehension ability and accurate textbase representation. Most of the studies described suggest that children of differing reading comprehension ability appear to possess the same literal memory for text and differ only in their ability to generate inferences. The exceptions to this are the studies by Oakhill (1984) and Cain, Oakhill, Barnes, and Bryant (2001) which found poor comprehenders to be impaired on their ability to answer questions requiring either literal or inferential information. Perfetti, Landi, and Oakhill (2005) argue that the production of inferences potentially reinforces the memory representation of literal propositions as children who generate inferences would be expected to form a more elaborate representation of the text and, therefore, to have a better literal memory for it.

Support for Perfetti et al.’s argument comes from research taking place within the “depth of processing” framework proposed by Craik and Lockhart (1972). “Depth of processing” refers to the extent of semantic or cognitive analysis accompanying the processing of stimuli. Craik and Lockhart argued that the strength of the memory trace of a stimulus depended on the “depth of processing” occurring during encoding. If associations are triggered and stimuli encoded with more elaboration, the memory trace is stronger. Early research in this area concentrated on the role of semantic processing in memory for individual words. For example, Schulman (1971) found that, in an unexpected memory test, recognition memory for words was better when participants had been asked to scan a word list for targets defined semantically (e.g. words denoting living things) rather than structurally (e.g. words containing the letter A). This demonstrated that memory of the words related to “depth of processing” during encoding. More recently, authors have extended the “depth of processing”

framework to explain findings from comprehension research, specifically the role of inferencing in memory for text. Friedman and Rickards (1981) presented college students with a text in which each paragraph was followed by a verbatim question, a paraphrase question or an inference question. Twenty-four hours later, students were tested on their recall of the material in the text using a sentence completion procedure. It was found that material contained in paragraphs which had been followed by inference questions was recalled better than material in paragraphs followed by paraphrase questions. This, in turn, was recalled better than material in paragraphs which had been followed by verbatim questions. Friedman and Rickards argued that the continuum of verbatim to paraphrase to inferential semantic processing represented a move from shallow to deep processing and that, in line with the “depth of processing” framework proposed by Craik and Lockhart, recall of text information was facilitated by increased semantic processing at encoding.

Given the argument presented by Perfetti et al. and the findings of research conducted within Craik and Lockhart’s “depth of processing” framework, it would be expected that children who were impaired on the ability to recall inferential information would also be impaired on the ability to recall literal information from text. Although this was found to be the case by Oakhill (1984) and Cain et al. (2001), it is not clear why these findings were not replicated in the other studies reported here (Cain & Oakhill, 1999; Oakhill, Cain, & Bryant, 2003).

Few studies investigating the role of inferencing in listening comprehension have been conducted. Recent research shows that the ability to generate inferences when material is presented aurally is correlated with listening comprehension in 4-, 6- and 8-year-olds (Kendeou, Bohn-Gettler, White, & van den Broek, 2008). Research has also suggested that, as in some of the reading comprehension studies discussed above, children who make the most elaborate inferences when listening possess the best memory for the text. Paris and Upton (1976) read short passages to children aged between 6 and 12. They asked questions requiring either memory of explicitly stated ideas or the ability to make knowledge-based inferences. They found that older children were more likely to generate inferences than younger children and that, as reading comprehension research has shown, this could not be explained entirely by their greater memory for the text. Furthermore, when a free recall task was

administered approximately 20 minutes after the test, the best predictor of number of idea units recalled was initial performance on inference questions requiring the integration of general knowledge with information from across several phrases or sentences. The authors concluded that, when listening, enhanced inferencing skill results in a more coherent representation of the text being formed and improved memory for literal information.

3.5. Exposure to print

Cunningham, Stanovich, and colleagues argue that the amount of time spent reading influences both word reading ability and reading comprehension in children and adults. They have developed a technique for assessing exposure to print, the Title Recognition Test (TRT). In this test, participants are presented with a list which includes real book titles and foils which sound like book titles. They are asked to indicate which of the titles they recognise. Guessing can be corrected for because of the inclusion of the foils. The number of titles correctly recognised, having corrected for guessing, is an index of the participant's reading experience.

Exposure to print, assessed using the TRT, has been shown to explain additional variance in orthographic processing after controlling for phonological processing in both adults (Stanovich & West, 1989), and children (Cunningham & Stanovich, 1990). Orthographic processing tasks assess, for example, the ability to discern the correctly spelled word from a pair of letter strings that sound the same (e.g. rume-room). The authors argue that their finding demonstrates that exposure to print leads to the development of the orthographic lexicon which in turn enhances word recognition ability.

Exposure to print also predicts reading comprehension ability directly. In their longitudinal study, Cipielewski and Stanovich (1992) found that performance on the TRT at age 11 predicted variance in reading comprehension after statistically controlling for decoding skill and reading comprehension at age 9. Cunningham and Stanovich (1997) conducted a longitudinal study of children over a 10 year period, from ages 7 to 17. They found that exposure to print at age 17 predicted unique

variance in reading comprehension at the same age having controlled for reading comprehension ability at age 7.

Exposure to print has been found to predict growth in skills, other than word recognition ability, which underpin reading comprehension ability. For example, Echols, West, Stanovich, and Zehr (1996) found that, after controlling for previous vocabulary knowledge, children's performance on a test of receptive vocabulary was predicted by the TRT administered a year earlier.

Importantly, exposure to print is also related to general knowledge and its development. Echols et al. (1996) found performance on the TRT to predict growth in general knowledge in children. Similarly, Stanovich and Cunningham (1993) found that exposure to print was a much better predictor than general cognitive ability of general knowledge in adults. The more prolific readers had a better general knowledge across a wide variety of domains. Knowledge itself is crucially important in reading comprehension (Perfetti, Marron, & Foltz, 1996). Without knowledge, comprehension is poor and the ability to learn new knowledge from text is impaired. It has been shown that, in adult readers, the representations of a text formed by "experts" in the subject are qualitatively different to the representations formed by "non-experts" (Long & Prat, 2002; Long, Wilson, Hurley, & Prat, 2006). High knowledge of the subject of a text leads to the construction of a more integrated and elaborated situation model of a text.

Despite the results reported above, not all studies have found a difference between good and poor reading comprehenders on a measure of exposure to print, especially when an Author Recognition Test (ART) is used instead of the TRT (Cain, Oakhill, & Bryant, 2000a; Ricketts, Nation, & Bishop, 2007).

There is some evidence to suggest that exposure to print leads to growth in listening comprehension. Hedrick and Cunningham (2002) found that listening comprehension at age 11 was predicted by performance on the TRT, having controlled for listening comprehension 2 years earlier. Senechal and le Fevre (2002) found that listening comprehension at age 7 was predicted by measures of storybook exposure during kindergarten having statistically controlled for parental education and phonological

awareness. Storybook exposure was assessed using two tests which were administered to parents of pre-school children. The first, a TRT, assessed recognition of titles of children's books and the second, an Author Recognition Test (ART), assessed recognition of authors of children's books. A relationship was also found between early exposure to storybooks and a composite measure of reading ability, incorporating both word recognition and reading comprehension, at age 9 after controlling for earlier reading ability. This was mediated, however, by early listening comprehension and vocabulary skills, a finding which the authors argue is consistent with the view that early exposure to storybooks enhances vocabulary and listening comprehension skills which in turn influence reading ability.

3.6. Working Memory

A huge amount of research has been conducted into the relationship between working memory (WM) and reading comprehension. The Construction-Integration model of comprehension specifies a role for a finite capacity WM in which processing occurs (Kintsch & Rawson, 2005). The exact nature of this WM system, however, has been the subject of much debate and various theories have been proposed (Long, Johns, & Morris, 2006). A discussion of the details of these conceptualisations is beyond the scope of this thesis. Nevertheless, two major theoretical models can be identified. The first is Baddeley's (1986) fractionated model in which the phonological loop and visuo-spatial sketchpad are responsible for the storage of information whilst the central executive co-ordinates storage and manipulates information. The second is a limited capacity unitary system in which storage and processing operations compete for resources (Daneman & Carpenter, 1980). Research into the role of WM in reading comprehension has generally taken place within the framework provided by this second model.

There is general agreement that "storage" capacity, sometimes referred to as short term memory (STM), does not differentiate good and poor adult comprehenders (Daneman & Carpenter, 1980; Daneman & Merickle, 1996), although some authors have found otherwise (Engle, Nations, & Cantor, 1990; LaPointe & Engle, 1990). STM is measured using simple span tasks such as digit or word span. In children, Engle, Carullo, and Collins (1991) found that the correlation between reading

comprehension and performance on a simple span task was the same as the correlation between reading comprehension and performance on a complex span task. However, these authors did not take into account the children's word recognition ability, so it may be that the relationship between word span and reading comprehension reflected differences in word reading ability. Certainly, most studies comparing good and poor reading comprehenders matched on word recognition ability have found no group differences in overall simple span (Cain, Oakhill, & Lemmon, 2004; Stothard & Hulme, 1992) or in patterns of performance on the simple span task (Oakhill, Yuill, & Parkin, 1986). A study in which good and poor comprehenders were matched on non-word reading ability did, however, find subtle group differences in performance on the simple span task (Nation, Adams, Bowyer-Crane, & Snowling, 1999). Whilst poor comprehenders recalled concrete words as well as good comprehenders, they were poorer when the recall task involved abstract words. The authors suggested that this finding reflected the poor comprehenders' underlying semantic problems. Cain (2006) did not replicate these findings when groups were matched on word recognition ability, possibly because, as mentioned earlier, this selection procedure may effectively "screen out" children with semantic impairments.

The lack of a relationship between simple storage capacity (STM) and reading comprehension led Daneman and Carpenter (1980) to devise a complex span task to measure working memory (WM), the ability to simultaneously store and process information. In the reading span task, participants read a series of sentences and are then asked to recall the final word of each sentence. Span is measured by the number of sentences for which final word recall is accurate. Daneman and Carpenter found that, in a population of college students, performance on the reading span task was correlated with reading comprehension. They argued that good readers could process linguistic material more efficiently than poor readers, so possessed greater functional storage capacity as less processing capacity was being used. Individual differences in complex WM span therefore reflected differences in processing efficiency. The correlation of performance on tasks tapping WM capacity and reading comprehension in adults is a very robust finding (Daneman & Merickle, 1996). However, a criticism of research into comprehension using span tasks involving linguistic material is that the WM measures themselves involve language comprehension. In other words, the

trivial conclusion could be drawn from such studies that sentence comprehension correlates with paragraph comprehension (Daneman & Merickle, 1996). Numerical complex span tasks, involving the manipulation and storage of numerical information, have, therefore, been developed to investigate whether the relationship of WM to reading comprehension extends to non-linguistic WM measures. In their meta-analysis of 77 studies involving more than 6000 participants, Daneman and Merickle (1996) found that complex span tasks involving numerical material, such as the operation span task of Turner and Engle (1989), do predict reading comprehension but not as well as those involving linguistic material. The authors suggested that comprehension is related to the efficiency with which symbolic computations rather than simply verbal computations can be carried out. Interestingly, however, Engle and colleagues interpret these findings differently. They argue that the relationship of reading comprehension and performance on numerical WM tasks indicates that WM tasks measure “general capacity” (Engle, Cantor, & Carullo, 1992). According to this view, individual differences arise because of differences in the quantity of resources available rather than because of the efficiency of certain processes.

There is now a large body of literature examining the relationship between reading comprehension and complex span in children. Stothard and Hulme (1992) compared 8-year-old good and poor comprehenders matched for word recognition ability on a listening span task in which children had to verify sentences and recall the final word of each. They were found not to differ on this verbal WM task. It has been argued that, as performance was generally low, floor effects may have been masking differences between the groups (Cain, Oakhill, & Lemmon, 2004). In general, most studies investigating the relationship between reading comprehension and performance on verbal complex span tasks in children have found that better comprehenders perform better on the WM measure. Cain (2006) found that 9-10 year-old good and poor comprehenders matched on vocabulary and word recognition ability differed significantly in a listening span task. Leather and Henry (1994) found that listening span made a significant unique contribution to reading comprehension after statistically controlling for simple span and phonological awareness. Similarly, Seigneuric, Ehrlich, Oakhill, and Yuill (2000) found that listening span predicted reading comprehension in French children after statistically controlling for word recognition ability and vocabulary. In their longitudinal study, Cain, Oakhill, and

Bryant (2004) found that, for children aged 8, 9 and 11, listening span predicted unique variance in reading comprehension having controlled for contributions of word recognition accuracy, sight vocabulary, receptive vocabulary and verbal IQ. Seigneuric and Ehrlich (2005) also conducted a longitudinal study with French children and found that listening span of 8-year-olds predicted their reading comprehension at age 9 after accounting for the autoregressive effect of reading comprehension.

The studies described above suggest that the relationship of WM with reading comprehension is not mediated entirely by verbal skills. Nevertheless, as explained earlier, a problem with the use of a linguistic measure of WM span is that it has been criticised as potentially giving the good comprehenders an advantage. To investigate whether good and poor comprehenders differed in performance on a measure of WM which was not itself dependent on comprehension skill, Yuill, Oakhill, and Parkin (1989) developed a numerical WM span task. In this task, children were presented with triplets of numbers which they had to read. They then had to recall the final digit from each triplet. The authors found that good comprehenders performed significantly better than poor comprehenders on this WM task and concluded that comprehension in children is related to a general ability to simultaneously store and process information, whether it is linguistic or not. However, the findings from other studies involving a numerical WM task have been ambiguous. Cain, Oakhill, and Bryant (2004) found that performance on the numerical WM task did correlate with reading comprehension for children at age 9 but not at ages 8 and 11, whilst Seigneuric et al. (2000) found that performance on the numerical WM span task did not contribute uniquely to reading comprehension once the contribution of a linguistic measure of WM span had been accounted for. Leather and Henry (1994) used a counting span task in which children had to count the dots on each of a set of white cards and then recall the number of dots on each card in order. Although count span correlated with reading comprehension, it did not make a unique contribution to reading comprehension after statistically controlling for simple span and phonological awareness. Similarly, Goff, Pratt, and Ong (2005) used a backward digit span task as a measure of WM span and found that this did not explain variance in reading comprehension after statistically controlling for word recognition ability, vocabulary and syntactic skills. Furthermore, Cain (2006) found that although good

comprehenders performed better than poor comprehenders on the counting span task, the difference between the groups was greater when their performance on a listening span task was measured.

It would seem, therefore, that whilst the relationship of reading comprehension with performance on tests of verbal WM span is well established in children, its relationship with numerical WM span is still debatable. Critics have argued that even those studies demonstrating a relationship of reading comprehension with numerical measures of WM cannot be taken as evidence that comprehension is dependent on a general ability to store and process information. These critics argue that performance on numerical tasks involving memory for digit names is itself verbally mediated (Swanson & Berninger, 1995; Nation, Adams, Bowyer-Crane, & Snowling, 1999). To determine whether or not comprehension ability in children is related to a general ability to simultaneously store and process information rather than the ability to simultaneously store and process verbal information only, spatial WM measures have been devised. Nation et al. (1999) used a complex odd-one-out task in which children were presented with three shapes and had to choose the odd-one-out, remembering its position for later recall. Children were also given a listening span task. Poor comprehenders matched with good comprehenders on non-word reading ability were found to perform less well on the listening span task. There was no difference, however, between the groups on the spatial span task. Similarly, Seigneuric et al. (2000) found that reading comprehension was not correlated with performance on a grid test in which children had to supply a missing dot to complete a line and remember the position of the line for later recall.

Interpretation of the results presented here is difficult. Evidence seems to suggest that reading comprehension ability is related to performance on linguistic WM span tasks. This is not surprising as it would be expected that poor comprehenders would have more difficulty with the processing demands of the tasks and would, therefore, have less functional storage capacity. Results from studies investigating the relationship of reading comprehension with WM tasks that do not use linguistic material are more inconsistent. There does not as yet appear to be sufficient evidence to conclude that poor comprehenders are impaired on a general ability to simultaneously store and

process information (Nation, Adams, Bowyer-Crane, & Snowling, 1999; Swanson & Berninger, 1995).

Very few studies have considered the relationship of either STM or WM with listening comprehension. The only study which has investigated the relationship between WM and listening comprehension in adults is the original Daneman and Carpenter (1980) study which found that the correlation of WM with listening comprehension was lower than that with reading comprehension. Also, listening comprehension correlated more highly with the listening span task than with the reading span task whilst the reverse was true for reading comprehension. As far as listening comprehension in children is concerned, Daneman and Blennerhasset (1984) found performance on a STM word span task correlated with listening comprehension in 3-5 year-olds, but failed to explain unique variance in listening comprehension once performance on a complex WM span task had been taken into account. Adams, Bourke, and Willis (1999) found that correlations between listening comprehension and measures of simple span were similar to the correlation between listening comprehension and a measure of complex WM span in 4-5 year-old children. Similarly, Florit, Roch, Altoe, and Levorato (2008) found that the listening comprehension of 4- and 5- year-old Italian children was independently predicted not only by a measure of complex WM span but also by a word span task, after statistically controlling for verbal IQ, and receptive vocabulary. Furthermore, the contributions of STM and WM to listening comprehension were similar. Florit et al. hypothesised that the importance of the measure of storage may reflect the greater storage demands of listening comprehension over reading comprehension. The relationship of listening comprehension with performance on a complex span task involving non-linguistic materials has not been assessed.

3.7. Non-Verbal Intelligence

Studies into reading comprehension in children often control for non-verbal intelligence. For example, in studies in which good and poor comprehenders are compared, children may be matched on non-verbal intelligence (Cain & Oakhill, 2006a; Nation & Snowling 1998b; Nation & Snowling, 1999). Similarly, when

regression techniques are used, non-verbal intelligence is often entered as a control variable (Nation & Snowling, 2004; Oakhill, Cain, & Bryant, 2003).

However, studies which have explored the role of non-verbal intelligence in reading comprehension more fully are limited and have not produced consistent results. Tiu, Thompson, and Lewis (2003) found that, in a sample of children aged 11, non-verbal intelligence explained additional variance in reading comprehension having controlled for listening comprehension, decoding ability and processing speed. Stanovich, Cunningham, and Feeman (1984), however, found that non-verbal intelligence was not related to reading comprehension, once decoding ability had been accounted for, in two groups of children aged 7 and 11.

Nation and Snowling (2002) compared good and poor reading comprehenders aged between 7 and 9 on their general cognitive ability, assessed using verbal tasks, non-verbal tasks and spatial tasks. Having been matched for chronological age and decoding ability, the two groups were found to differ on verbal skills and non-verbal reasoning but not on spatial ability. Furthermore, a small percentage of the sample of poor comprehenders was found to have, overall, very poor general cognitive ability. The authors argued that, for some children, poor general cognitive ability, rather than weaknesses specific to the verbal domain, may be a contributing cause to their reading comprehension difficulties. In a later study, however, Cain and Oakhill (2006a) found that, at age 8, good and poor reading comprehenders matched on word recognition accuracy did not differ on a measure of non-verbal intelligence.

3.8. Summary and introduction to Study 1

As should by now be apparent, a great deal is known about the component skills involved in reading comprehension in both children and adults. Even so, many of the studies discussed in this section have considered a small number of variables at a time, so little is known about the relative importance of the predictors identified. Studies investigating the contributions of several variables to reading comprehension have been conducted but consistent results have not been obtained. For example, Cain, Oakhill, and Bryant (2003) report that a measure of complex WM span does explain additional variance in reading comprehension having accounted for contributions of word- and sentence-level language skills, a finding not replicated by Goff, Pratt, and Ong (2005). Nevertheless, these studies are important in attempting to explore the relative contributions of different variables to reading comprehension.

By comparison, very little is known about the skills underpinning listening comprehension. The few studies that have been conducted have investigated, almost exclusively, the component skills of listening comprehension in pre-schoolers (Adams, Bourke, & Willis, 1999; Florit, Roch, Altoe, & Levorato, 2008).

In Section 2.5.5, it was argued that an investigation is necessary into the assumption, formalised in the “Simple View of Reading”, that listening comprehension and reading comprehension are underpinned by the same component skills and processes, once the role of word recognition skills in reading comprehension is accounted for. The material reviewed in this chapter has shown that, currently, very much more is known about the cognitive-linguistic skills underpinning reading comprehension than about those underpinning listening comprehension. Some authors have attempted to make comparisons between results obtained from the studies of listening comprehension and reading comprehension. For example, Florit et al. (2008) speculate that the importance of a measure of the simple storage of information, forward word span, in listening comprehension but not reading comprehension results from the extra storage demands involved in listening comprehension. However, as pre-schoolers, the children in Florit et al.’s study were only assessed on listening

comprehension. Essentially, the authors are comparing their findings with the findings of studies involving older children.

Other studies have compared the processes involved in reading comprehension and listening comprehension in a more rigorous manner by assessing children on comprehension in each modality as well as on possible component skills. The contributions of predictor variables to comprehension in each modality can then be compared. This approach was taken by Hagtvet (2003) and by Burgoyne et al. (2009). These studies are important in illuminating some of the differences between reading comprehension and listening comprehension. However, in both only a small number of predictor variables were considered. Burgoyne et al. investigated the roles of expressive and receptive vocabulary in comprehension in the two modalities, whilst Hagtvet considered semantic, syntactic and phonological skills.

The research presented in Study 1 was conducted to ascertain whether or not reading comprehension and listening comprehension involve the same component skills. Specifically, the first aim of the study was to identify and compare unique predictors of reading comprehension and of listening comprehension. To meet this aim, some of the approaches outlined above were combined in a novel manner. Whilst Cain et al. (2003) compared the contributions of a wide range of variables to reading comprehension skill and to word recognition ability, Hagtvet (2003) and Burgoyne et al. (2009) compared the contribution of a smaller number of variables to reading comprehension and to listening comprehension. In Study 1, these approaches were combined by comparing the contributions of the wide range of variables described in this chapter to reading comprehension and to listening comprehension.

The second aim of the study was to test the Simple View of Reading directly by ascertaining whether any of the predictor variables described in this chapter continued to predict unique and significant variance in comprehension in a given modality after controlling for comprehension in the other modality. As described in Section 2.5.3, a number of previous studies have attempted to identify variables which predict additional variance in reading comprehension having controlled for listening comprehension and word recognition or decoding skills (e.g. Connors, 2009; Joshi & Aaron, 2000; Ouellette & Beers, 2010). One study has also explored whether

additional variance in listening comprehension can be explained by a predictor variable, vocabulary, having controlled for reading comprehension (Braze et al., 2007). Braze et al., however, considered the comprehension of sentences only, rather than extended discourse. The demands made by the comprehension of discourse in each modality, which are over and above those shared with comprehension in the other modality, have not previously been investigated.

Chapter 4.

Study 1: Are listening comprehension and reading comprehension underpinned by the same cognitive-linguistic skills? An investigation using standardised tests.

4.1. Summary of aims

The first aim of the study was to identify the skills, out of a selection of language, memory and intelligence variables, which made independent contributions to reading comprehension in a relatively large representative sample of children, and to compare these with the skills making independent contributions to listening comprehension. If, as assumed by the Simple View of Reading, exactly the same cognitive-linguistic processes underlie reading comprehension and listening comprehension, it would be expected that comprehension in each modality would be uniquely explained by the same predictor variables with additional variance in reading comprehension being explained by word recognition ability. The second aim was to identify which cognitive-linguistic skills continued to predict unique and significant variance in comprehension in each modality after controlling for comprehension in the other modality. If the Simple View of Reading is correct, reading comprehension should be predicted by word recognition ability only, once listening comprehension is controlled for. Furthermore, variance in listening comprehension would not be expected to be predicted by any other cognitive-linguistic skills after controlling for reading comprehension.

4.2. Method

Design

Two standard regression analyses were conducted with reading comprehension, as measured by the Neale Analysis of Reading Ability (NARA-II), and listening comprehension, as measured by the Listening Comprehension Test Series, as criterion variables. Predictor variables entered into each regression were word recognition accuracy (NARA-II), age, vocabulary (BPVS-II), syntactic ability (TROG-2),

semantic awareness (TOWK), exposure to print (TRT), inferencing ability, non-verbal intelligence (Raven's CPM), short-term memory (WMTB-C digit recall) and working memory (WMTB-C count recall). Comprehension scores in each modality were then regressed onto comprehension scores in the other modality and the residuals saved. Residualised measures of comprehension in each modality were then entered as criterion variables in two further standard regression analyses. Predictor variables were the same as those used previously.

Participants

Three Calderdale primary schools were involved in the study. According to the most recently published OFSTED reports, all three schools were attended predominantly by White British children with a very small number of children from other ethnic backgrounds. The schools had an average or lower than average number of children on free school meals, or on the school register of Special Educational Needs. They all followed the National Curriculum.

The study was designed to investigate the predictors of reading comprehension and listening comprehension across a range of abilities, typical of those that might be found in a primary school classroom. Thus, the sample was relatively unselected: whole year groups were assessed, as in previous studies (Goff, Pratt, & Ong, 2005; Leather & Henry, 1994; Seigneure & Ehrlich, 2005). In the present study, the only children who were not included were those with a statement of Special Educational Needs due to learning difficulties. Children who needed to wear glasses to correct their eyesight were asked to do so whilst being assessed on any measure using visual stimuli.

Pupils were assessed in Year 3 or Year 4. Studies into children's reading comprehension tend to involve pupils in these year-groups as, by this age, most have mastered basic word recognition skills and their reading comprehension abilities can be assessed (e.g. Cain & Oakhill, 1999; Nation & Snowling, 1998b; Oakhill, Cain, & Bryant, 2003; Ricketts, Nation, & Bishop, 2007). In the current study, testing began in the school summer term with children in Year 3. Due to the fact that only 59 children had been assessed before the school summer holidays, testing continued in the autumn term. Children in Year 4 were assessed to keep the age range of the participants in the

study as narrow as possible. All testing was completed by the end of the children's first term in Year 4. The final sample included 126 children, 59 of whom were in Year 3 at the time of testing, whilst 67 were in Year 4. The age range of the children at start of testing was 7 years, 9 months to 9 years, 2 months.

Study variables and measures

Each child was administered a battery of tests. This included tests of reading comprehension and listening comprehension as well as tests of the variables described in Chapter 3. All variables assessed in the test battery are listed below and a description of the tests used is given.

Reading Comprehension and Reading Accuracy

These were assessed using the Neale Analysis of Reading Ability – Second Revised British Edition, Form I (NARA-II: Neale, 1997). The Neale test has been used extensively, over several decades, in reading comprehension research undertaken with children in the UK. Originally written in 1966, it was revised in 1989 and subsequently re-standardised in 1997. Thus, whilst in older studies the original NARA was used (Oakhill, 1982, 1984), later studies have made use of one of the revised editions (Cain, 2006; Nation, Clarke, & Snowling, 2002; Oakhill, Cain, & Bryant, 2003; Ricketts, Nation, & Bishop, 2007; Spooner, Gathercole, & Baddeley, 2006).

The test consists of a series of short passages of increasing length and complexity. Children are tested individually and have to read each passage aloud. Any words that they are unable to read correctly are supplied by the experimenter in such a way that reading fluency is maintained. A record is made of the number of words that a child is unable to read by themselves and this number is used to calculate a reading accuracy score. At the end of each passage, children are asked questions to assess their understanding of the passage. They are able to refer to the text to help them answer these questions. According to the test manual, questions assess “understanding of the main ideas of the narrative, the sequence of events and other details, plus some limited inference” (Neale, 1997, p.3). The test manual gives clear guidelines regarding acceptable answers. Testing stops when a child makes 16 or more reading errors on any of the passages 1-5, or 20 errors on passage 6. Thus, the number of

comprehension questions attempted by a child is constrained by their word reading ability.

For each child, the NARA-II gives a score for word reading ability in context, based on the number of reading errors made, and a score for reading comprehension, based on the number of questions about the passages that are answered correctly. The NARA-II can also be used to measure rate of reading. However, as mentioned previously, most research has not found a relationship between rate of reading and comprehension, once word recognition skills have been taken into account (Goff, Pratt, & Ong, 2005; Stothard & Hulme, 1996). Therefore, this was not assessed in this study.

The test manual reports reliability estimates (Cronbach's alpha) for the Accuracy measure of .81 for children aged 6.00-7.11 and .87 for children aged 8.00-9.11. For the Comprehension measure, reliability is given as .94 for children aged 6.00-7.11 and .95 for children aged 8.00-9.11.

Listening Comprehension

This was assessed using the Listening Comprehension Test Series (Level C) (LCTS: Hagues, Siddiqui, & Merwood, 1999). This standardised test is designed for use by classroom teachers to assess the comprehension skills of pupils listening to extended pieces of text. No reading or writing is involved in the test.

In the Listening Comprehension Test Series, children hear four passages presented on a tape. Unlike the NARA-II, these passages do not increase in difficulty. Different versions of the test are available for different age-groups, so each passage is deemed to be of an appropriate level of difficulty for the age of the child. Level C is standardised for children aged 8:00-10:03 (10 years, 3 months) and was chosen as the most appropriate version to use in this study, even though a small number of the children assessed in Year 3 (6% of total sample) were younger than 8:00 years, the youngest child being 7:09 (7 years, 9 months) at time of test.

The passages span a range of listening contexts. In the version administered in this study, passages include a story, a poem, a discussion between a librarian and some

children and a description of a school given by a child as if they are conducting somebody around the building. Children hear each passage twice. After listening to a passage on the tape for the first time, children hear questions assessing their understanding of it. Questions are read aloud to them by the experimenter but are not attempted at this point. Children then listen to the passage for a second time. When the passage has ended, the experimenter once again reads the questions and the children attempt to answer them. The test authors argue that giving children the opportunity to listen to each passage twice, ensuring that they know what to listen for on the second hearing, closely replicates real-life situations by providing a context and focus for the children's listening.

Some questions require the children to circle the correct picture out of a choice of four, whilst others require them to decide whether the statement read out by the experimenter is true or false. One point is administered for each correct answer of the first type but, in most instances, for a point to be awarded for the true/false questions, two consecutive questions need to be answered correctly. Like the NARA-II, the questions are designed to assess a range of aspects of language comprehension, "the retrieval of specific information, the drawing of simple and more complex inferences and the synthesis of the material heard" (Hagues, Siddiqui, & Merwood, 1999, p.7).

The test is designed to be administered to whole classes and lasts about an hour. The test manual suggests that it can be administered in one session or over two separate sessions. In the current study, it was decided to administer the test over two separate sessions to avoid children becoming tired and losing concentration.

The test manual reports a reliability estimate (Kuder-Richardson 20) for the Listening Comprehension Test Series (Level C) of .83.

Receptive Vocabulary

This was assessed using the second edition of the standardised British Picture Vocabulary Scale (BPVS-II: Dunn, Dunn, Whetton, & Burley, 1997). This test has previously been used to assess receptive vocabulary in UK studies of children's comprehension (Cain, 2006; Oakhill, Cain, & Bryant, 2003; Muter, Hulme, Snowling, & Stevenson, 2004; Spooner, Gathercole, & Baddeley, 2006).

The test employs a multiple choice format. For each test item, children see four pictures and have to indicate which picture illustrates the meaning of a word presented aurally by the examiner. Test items are grouped in sets of 12. There are 14 sets and each set is more difficult than the preceding one. Testing ceases at the set on which eight or more errors are made.

In the current study, children were tested individually according to the guidelines given in the manual.

Reliability estimates of the BPVS–II reported in the manual are high. Corrected split-half reliabilities for individual age-groups are reported, the median being .86 which, it is argued, gives the best overall measure of the test’s reliability.

Semantic Awareness

This was assessed using the Word Definitions subscale of the Test of Word Knowledge (TOWK: Wiig & Secord, 1992). This test has been used by researchers in both the UK and the USA (Nation & Snowling, 1998b; Ouellette, 2006).

For each test item, children hear a word and are shown it in print form. They then have to talk about the meaning of the word. The scoring system used is based on the principle that a mature definition of a word includes the semantic category to which a word belongs and at least two unique semantic features. Points are awarded only for definitions including at least two of these components. A definition providing all three of these components is given a score of 2, and a definition containing two of these components is given a score of 1, whilst a definition giving just one component or none at all is given a score of 0. For example, for the word “teacher”, the definition “a person who helps you learn, they read stories” would be awarded 2 points, whilst “a person who teaches” would be awarded 1 point only. The definition “a person” would receive no points.

According to the instructions given in the test manual, testing should cease at the point at which five consecutive scores of 0 are obtained. However, the test is standardised for a North American population and the difficulty of the items does not

appear to increase in a uniform manner for a British sample. Specifically, the words “jug” and “bib” (items 19 and 20), are presented after the words “architect”, “tournament” and “chaperone” (items 16, 17 and 18 respectively). Thus, all children were administered the first 20 items and testing stopped when they had received five consecutive scores of 0 after this point. The mean scores on each item are given in Appendix 2 and justify this approach. In practice, few children (12% of whole sample) would have been excluded from attempting items 19 and 20 if the guidelines given in the manual had been followed.

Children were tested individually. Children’s definitions were tape recorded for later transcription and scoring.

For the modified version of the test being used in this study, Cronbach’s alpha was calculated as .78.

Syntactic Ability

This was assessed using the Test for Reception of Grammar, Version 2 (TROG-2: Bishop, 2003). Version 1 of this standardised test has previously been used to assess receptive grammar skills in studies of children’s comprehension (Goff, Pratt, & Ong, 2005; Stothard & Hulme, 1992) and has been used in a modified form by authors concerned about ceiling effects on some items (Nation, Clarke, Marshall, & Durand, 2004; Oakhill, Cain, & Bryant, 2003).

The test employs a multiple choice format, similar to the BPVS-II. For each test item, children see four pictures and have to indicate which picture illustrates the meaning of a sentence presented aurally by the examiner. There are 80 items altogether arranged in 20 blocks of 4. Each block tests understanding of a specific grammatical construct and is more difficult than the preceding block. According to the test manual, a block is failed if the meaning of at least one of the four items is incorrectly identified. Testing ceases when five consecutive blocks are failed.

In the current study, children were tested individually according to the guidelines given in the manual.

Split-half reliability for the TROG-2 is reported in the manual as being .88.

Off-line Inference Generation

This was assessed using a task adapted from that devised by Oakhill (1984). This task was selected because it specifically assesses children's ability to make knowledge-based, gap-filling inferences.¹

Oakhill's (1984) test was itself adapted from materials devised by Paris and Upton (1976). Paris and Upton read short passages to children and asked them to answer questions requiring either memory of explicitly stated ideas or the ability to make knowledge-based inferences. This test assessed the ability to generate different sorts of inference. Some inferences were "lexical", constrained primarily by single words, whilst others were "contextual", requiring sentential or intersentential information. Both types, however, required the integration of information stated explicitly in the text with general knowledge. Forced-choice questions were used; children had to respond "yes" or "no" to the statements presented to them. Oakhill (1984) adapted the materials used by Paris and Upton so that passages were read by the children themselves. Also, questions were no longer of the yes/no type but were open-ended, requiring children to formulate their own answers. Cain and Oakhill (2006) have argued that open-ended questions assess children's ability to generate inferences much more accurately than forced choice questions as it has been shown that recognition of an inference does not necessarily demonstrate that the inference was generated at encoding (Corbett & Doshier, 1978).

In the current study, as children with a wide range of word reading abilities were included, it was important that performance on the inferencing test was not dependent on reading ability. Thus, Oakhill's task was adapted so that children listened to the passages read by the experimenter as in the original Paris and Upton (1976) study. Only questions requiring the ability to make knowledge-based inferences were asked. Questions requiring memory for information stated explicitly in the text were not

¹ More recent tests constructed by Oakhill and colleagues have assessed children's text-connecting inferencing abilities as well as their gap-filling inferencing skills (Cain & Oakhill, 1999; Oakhill, Cain, & Bryant, 2003). However, as the role of text-connecting inferencing in comprehension has been questioned (Spooner, Gathercole, & Baddeley, 2006), it was decided that only the generation of knowledge-based inferences would be assessed in this study. Therefore, the older test was chosen.

included, as research investigating the role of explicit memory in reading comprehension has produced mixed results, as explained previously (Cain & Oakhill, 1999; Oakhill, Cain, & Bryant, 2003). Questions were open-ended and required children to formulate their own responses.

Children were tested individually. They listened to four short passages which were each followed by four questions. Their responses to the questions were noted for later scoring. The stories were presented to all children in the same order and the questions for each story were always asked in the same order. An example of a story from the test is shown in Table 1 and a copy of the full test is given in Appendix 3.

Table 1: Example story from the test of inferencing ability

Tim and the Biscuit Tin

Tim waited until he was alone in the house. The only sound he could hear was his father's axe on the logs in the shed. Tim looked in all the rooms again, to make sure his mother was not there. Then he pushed a chair over to the sink which was full of dishes. By climbing onto the edge of the sink, he could just reach the biscuit tin. The tin was behind the sugar. Tim stretched until his fingers could lift up the lid. Just as he reached inside, the door swung open and there stood his little sister.

1. Why did Tim want to be alone in the house? *So he could steal/get/eat the biscuits. So he wouldn't be caught/told off*
2. What room was Tim in? *Kitchen*
3. What was Tim's father doing? *Chopping/cutting logs/wood*
4. How did Tim climb onto the sink? *By pushing a chair over to the sink and standing on it. By using/standing on a chair*

Oakhill (1984) herself acknowledges the difficulty in scoring a test when questions are of an open-ended rather than a forced choice nature. It is not always easy to determine what constitutes a correct answer. For this study, fairly stringent criteria were applied. Children needed to demonstrate that they had spontaneously gone beyond the information that they had been given to make an appropriate inference in order to be scored as giving a correct response. A list of responses accepted as correct

is given in Appendix 3. A second rater scored a third of the sample using this rubric and inter-rater agreement was found to be very high (agreement on 97.5% of items).

Split-half reliability for the test (Spearman-Brown correction) was calculated as .74. A limitation with the measure of off-line inferencing adopted in this study, highlighted by Cain and Oakhill (1999) is that performance on the test relies on children possessing the relevant general knowledge. If the general knowledge required to draw a particular inference is not possessed by a child then the inference cannot be made. Cain and Oakhill (1999) overcame this problem by asking children questions assessing whether or not they possessed the required general knowledge, if they failed to answer an inferencing question correctly. However, this approach has only been taken in situations where the children have read the stories themselves and have the text available to them. It is arguably harder to do this and potentially more disruptive to the administration of the test when children have listened to the stories. Vocabulary, however, is highly correlated with general knowledge (Cunningham & Stanovich, 1991; Ransby & Swanson, 2003), as is print exposure (Stanovich & Cunningham, 1993). In this study, therefore, the role of general knowledge in knowledge-based inferencing is effectively controlled for by the inclusion of these other variables. Any relationship between knowledge-based inferencing and comprehension, independent of the relationship between vocabulary and comprehension, is assumed to reflect the tendency to spontaneously generate inferences rather than higher levels of general knowledge.

Exposure to Print

This was assessed using an adapted version of the Title Recognition Test (TRT: Cunningham & Stanovich, 1990). This has been found to be a good measure of print exposure in children. It uses a signal detection logic in which real titles are embedded among foils (titles that sound plausible but are in fact not real). The use of guessing as a strategy can therefore be corrected for.

In their original version of the TRT, Cunningham and Stanovich used 39 items, 25 real titles and 14 foils. However, other authors have adapted this test so that the probability of selection of a foil is equal to the probability of selection of a real title

by including equal numbers of target items and foils (Goff, Pratt, & Ong, 2005; Ricketts, Nation, & Bishop, 2007). This was the approach used in the current study.

This test was developed specifically for this study. A list of 20 popular children's books was compiled. Books were chosen based on visits to bookshops and libraries. The selected titles covered a range of types of books. A list of 20 foils was also created. These foils were randomly interspersed amongst the real titles. A list of the titles, both real and foils, is given in Appendix 1, along with the percentage correct recognition for each item.

Children were assessed individually and the titles read out to them to control for different word reading abilities. Instructions given were as follows:

"I'm going to read you a list of names of books and I want you to tell me if you've heard of them. It's important that you don't guess and only say yes to the books that you know are real because some of these names are made up and are not real books at all"

Because there were equal numbers of target items and foils, the TRT was scored by deducting the number of false alarms from the number of hits.

Appendix 1 shows the percentage recognition for each item. As can be seen, a few items have poor psychometric properties, demonstrating ceiling or floor effects. Furthermore, the reliability estimate (Cronbach's alpha) was 0.57 for target items, which is low. In the original TRT (Cunningham & Stanovich, 1990), Cronbach's alpha is reported as .81. Therefore, in the current study, results involving the TRT must be interpreted with care because of the relatively low reliability of the test used.

Working Memory/ Short-term Memory

These variables were assessed using subtests from the standardised Working Memory Test Battery for Children (WMTB-C: Pickering & Gathercole, 2001). Children were tested on a simple span task, the digit recall scale, to assess short-term memory, the ability to store information. They were tested on a complex span task, the count recall scale, to assess working memory, the ability to simultaneously store and process

information. Numerical span tasks were chosen because, as explained earlier, performance on linguistic measures of working memory span may, to some extent, be dependent on language comprehension ability. The subtests of the WMTB-C chosen for the current study have been used in previous research into children's comprehension (Cain, 2006).

In the digit recall task, sequences of digits are presented aurally to the child at the rate of one digit per second. After each sequence, children have to recall the digits in the correct order. Children are given practice trials involving recall of one, two and three digits. Experimental trials begin with sequences of three digits. Sequences range from one to nine digits in length and are arranged in blocks such that, for each sequence length, there are six trials. However, once a child has successfully completed four trials at any given sequence length, they move to the next block and those trials left unattempted are credited as correct. Testing ceases when three or more errors are made within a block.

In the count recall task, children are presented with a series of white cards, one at a time. On each card there are several coloured dots which the child has to count. The number of dots ranges from four to seven. After counting the dots on the last card, children have to recall the number of dots on each card in the order in which they were presented. Children are given practice trials involving one, two or three cards and experimental trials begin with the two-card sets and continue with progressively longer sets. Sets range from one to seven cards in length and are arranged in blocks such that, for each set length, there are six trials. When a child has successfully completed four trials at a certain set length, they move to the next block, omitted trials being credited as correct. When three or more errors are made within a block, testing stops.

According to the manual, these subtests can be scored in one of two ways. Span score is the longest number of digits/ cards at which recall is successful. In other words, it is the length of the sequence corresponding to the penultimate block of trials administered. However, the test also allows for a Trials Correct score to be calculated. This is the overall number of trials correctly recalled up to the point at which testing stops. Following Leather and Henry (1994), this measure was adopted in the current

study as it is a more sensitive indication of ability, enabling small differences between individuals to be recorded.

Tests were administered to children individually in accordance with guidance given in the test manual.

Test-retest reliability for the digit recall subtest of the WMTB-C is given in the manual as .81 for Years 1 and 2, and .82 for Years 5 and 6. For the count recall subtest, test-retest reliability is given as .74 for Years 1 and 2, and .48 for Years 5 and 6.

Non-verbal IQ

This was assessed using Raven's Coloured Progressive Matrices (CPM: Raven, Raven, & Court, 1998). This test is designed specifically to assess the perceptual reasoning processes of children under the age of 11 and has previously been used in studies of children's comprehension (Stanovich, Cunningham, & Feeman, 1984).

Raven's CPM consists of 36 items organised as three sets of 12. Each item is presented as a coloured illustration with a missing section. The children's task is to choose the figure from a choice of 6 which, when inserted in the picture, successfully completes the pattern. Little verbal explanation is necessary. The children are helped with the first two problems of the first set if necessary to ensure that they know what is expected of them. Children complete all items.

In the current study, children were tested individually according to the guidelines given in the manual.

In several studies reported in the manual, split-half reliability coefficients have been found to be high, ranging between .82 and .94, although lower estimates have sometimes been reported for children younger than 6.

Procedure

Most of the assessments were carried out with children individually in a quiet area of the school. Because of the large number of tests, they were administered over three sessions each lasting approximately 30/40 minutes. The individually administered tests were piloted on six children in Year 2. This pilot demonstrated that children understood the instructions for each task and that the testing sessions were short enough for children to maintain concentration.

In the first session, children were assessed using the TROG-2 and the NARA-II. In the second session, children were assessed using the WMTB-C, the TOWK and the BPVS-II. In the third session, the off-line inference generation task and Raven's CPM were administered. For most children, sessions were separated by between one and two weeks.

The Listening Comprehension Test Series (Level C) is a whole class test which can be administered either in one session of between 45-60 minutes length or over two separate sessions. The test was piloted on six children in Year 2. This revealed that some children found it difficult to concentrate for longer than half an hour. It was decided, therefore, that in Study 1 the test would be administered to whole classes over two half-hour sessions, normally separated by about a week. The first whole class session was always held after each child had completed their first individual session, so that all the children felt at ease with the researcher.

4.3. Results

Sample size

An *a priori* power analysis was conducted. A medium-sized relationship between the predictors and the criterion variable was assumed ($f^2 = .15$), as was an alpha level of .05 and a beta level of .20. Results suggested that a sample size of 118 was required to test the multiple correlation. The sample size of 126 met these minimum requirements. Furthermore, the sample size of 126 nearly met the minimum requirements given by Tabachnick and Fidell (2007) for testing the multiple

correlation ($N > 50+80$), and met their requirements for testing individual predictors ($N > 104 + 10$).

There was a very small amount of missing data (0.07%). This was dealt with by excluding cases listwise from the analysis.

Data preparation

The data collected was analysed using regression techniques. However, in order to use regression techniques, certain assumptions have to be met. The following screening procedures were used to ensure that these assumptions had not been violated.

Investigation of univariate distributions

Regression techniques are sensitive to outliers and non-normal distributions (Tabachnick & Fidell, 2007). Thus, initial analyses were conducted to identify univariate outliers and atypical distributions.

Following Tabachnick and Fidell (2007), outliers were considered to be those cases with scores on one or more variables that were greater than 3.29 standard deviations above or below the mean. Only a very small percentage of cases were identified as outliers (2.38%). It was decided that outlying scores would be retained for analysis as visual inspection of histograms and expected normal probability plots for each variable suggested little deviation from normality. Values for skew and kurtosis were obtained and were found to be between +1 and -1 for each variable. The decision to retain the outlying scores was justified when further screening revealed no multivariate outliers (see below). Furthermore, when each regression analysis was run, Cook's distance was obtained and it was confirmed that no cases were exerting an undue influence on the results (see below).

Investigation of multivariate distributions

Multivariate outliers, cases with an unusual combination of scores on two or more variables, were investigated by obtaining the Mahalanobis distance for each case and evaluating it using the chi-square distribution. This was done by running an initial regression analysis in which the case number was specified as DV whilst the 10 predictor variables were specified as IVs. Mahalanobis distances were requested.

With the use of a $p < .001$ criterion, none of the cases had a value of critical chi square in excess of 29.588 (10 IV's), i.e. no multivariate outliers were detected.

Investigation of residuals

When each of the regression analyses reported in this chapter was carried out, the residuals were examined.

Outliers in the solution, cases for which scores on the DV were poorly fit by the regression equation, were requested. Such cases were identified as those for which the standardised residual was greater than +3 or lower than -3. Identification of such cases is important as their inclusion can lower the multiple correlation. However, in the regression analyses reported below, no outliers were identified.

Examination of the residuals also allowed the assumptions of normality, linearity and homoscedasticity of residuals to be tested. Normal probability plots and scatterplots of predicted values of the DV against standardised residuals were obtained to examine whether or not there were major deviations from normality, linearity and homoscedasticity. No deviations were identified.

Additionally, during the regression analyses, cases exerting excessive influence on the results were also screened for. Cook's distance was requested for each case as part of the regression output. No cases were found to possess a Cook's distance greater than 1 in any of the regression analyses reported. Thus, no case appeared to exert undue influence on the results.

Descriptive statistics

Because ages of participants were noted, age could be independently controlled for in the regression analyses that follow. Thus, for the standardised tests that were used, it was unnecessary to convert raw scores to standardised scores. Instead, raw scores were used for all measures. Table 2 shows the descriptive statistics obtained for each measure using these raw scores.

Table 2: Means and standard deviations for all Study 1 measures (raw scores)

Measure	N	Mean	SD	Skew	Kurtosis
NARA-II accuracy	126	54.73	22.42	-.046	-.782
Age	126	100.80	3.54	-.028	-.144
BPVS-II	126	89.10	12.97	-.239	.562
TROG-2	126	14.13	3.52	-.634	.007
TOWK	126	22.87	6.31	.028	-.727
TRT	126	8.46	2.86	-.043	.518
Inferencing	126	9.11	2.84	-.303	-.439
Raven's CPM	126	28.50	4.69	-.773	-.061
WMTB-C digit recall	126	27.03	3.87	.503	.913
WMTB-C count recall	126	17.48	4.27	.315	.653
NARA-II comprehension	126	19.90	8.21	-.176	-.555
LCTS comprehension	125	23.76	5.99	-.357	-.605

Key: NARA-II accuracy – Word recognition from the Neale Analysis of Reading Ability (Second Revised Edition); Age – chronological age in months; BPVS-II – British Picture Vocabulary Scales (Second Edition); TROG-2 – Test for Reception of Grammar (Version 2); TOWK – Test of Word Knowledge (Word Definitions sub-scale); TRT – Title Recognition Test; Inferencing – Correct inference responses; Raven's CPM – Raven's Coloured Progressive Matrices (1998 Edition); WMTB-C digit recall – Working Memory Test Battery for Children, simple span task; WMTB-C count recall – Working Memory Test Battery for Children, complex span task; NARA-II comprehension – Reading comprehension from the Neale Analysis of Reading Ability (Second Revised Edition); LCTS comprehension – Listening comprehension from the Listening Comprehension Test Series.

Regression analyses

To ascertain which variables explained unique variance in comprehension in each modality, two standard multiple regression analyses were conducted, with reading comprehension and listening comprehension as the criterion variables.

Bivariate correlations (two-tailed) between the variables were firstly obtained to check for evidence of multicollinearity between the predictor variables. Results are shown in Table 3. Because of the large number of correlations, a Bonferroni correction was applied, giving a criterion for significance of $p < .00076$ ($.05/66$).

Table 3: Correlations between Study 1 measures

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1.LCTS comprehension	-	.665**	.552**	.234	.714**	.609**	.641**	.484**	.554**	.604**	.357**	.378**
2.NARA-II comprehension		-	.887**	.224	.684**	.626**	.687**	.555**	.293*	.531**	.495**	.415**
3.NARA-II accuracy			-	.228	.581**	.507**	.632**	.550**	.200	.420**	.420**	.354**
4.Age				-	.315*	.203	.330*	.098	.048	.192	.120	.258
5.BPVS-II					-	.577**	.682**	.425**	.512**	.557**	.446**	.392**
6.TROG-2						-	.614**	.433**	.452**	.620**	.506**	.393**
7.TOWK							-	.473**	.433**	.504**	.376**	.394**
8.TRT								-	.271	.389**	.387**	.306*
9.Inferencing									-	.452**	.115	.347**
10.Raven's CPM										-	.420**	.438**
11.WMTB-C digit recall											-	.358**
12.WMTB-C count recall												-

** $p < .0001$, * $p < .001$

Key: NARA-II accuracy – Word recognition from the Neale Analysis of Reading Ability (Second Revised Edition); Age – chronological age in months; BPVS-II – British Picture Vocabulary Scales (Second Edition); TROG-2 – Test for Reception of Grammar (Version 2); TOWK – Test of Word Knowledge (Word Definitions sub-scale); TRT – Title Recognition Test; Inferencing – Correct inference responses; Raven's CPM – Raven's Coloured Progressive Matrices (1998 Edition); WMTB-C digit recall – Working Memory Test Battery for Children, simple span task; WMTB-C count recall – Working Memory Test Battery for Children, complex span task; NARA-II comprehension – Reading comprehension from the Neale Analysis of Reading Ability (Second Revised Edition); LCTS comprehension – Listening comprehension from the Listening Comprehension Test Series.

Investigation of the correlation matrix shown in Table 2 suggested no evidence of multicollinearity between the predictor variables. None of the correlation coefficients for relationships between predictor variables exceeded the value of .80.

Results of a standard multiple regression analysis with NARA-II reading comprehension as the criterion variable are shown in Table 4.

Table 4: Standard regression analysis with NARA-II reading comprehension as DV and language, memory and intelligence variables as predictors

Variable	B	SE B	Beta	t	p	sr ² (unique)
NARA-II accuracy	.239	.019	.652***	12.353	<.001	.193
Age	-.104	.091	-.045	-1.136	.258	
BPVS-II	.106	.037	.167**	2.888	.005	.010
TROG-2	.287	.127	.123*	2.260	.026	.006
TOWK	.078	.075	.060	1.050	.296	
TRT	.053	.129	.018	.409	.683	
Inferencing	-.134	.138	-.046	-.974	.332	
Raven's CPM	.090	.087	.051	1.034	.303	
WMTB-C digit recall	.067	.097	.031	.688	.493	
WMTB-C count recall	.068	.082	.036	.833	.406	

*** $p < .001$, ** $p < .01$, * $p < .05$

R for regression was significantly different to zero, $F(10,115) = 67.772$, $p < .001$, with R^2 at .855 and adjusted R^2 at .842.

The linear model accounted for a substantial amount of variance (84%) in reading comprehension. Only three regression coefficients differed significantly from zero: word recognition (NARA-II accuracy), vocabulary (BPVS-II) and syntactic abilities (TROG-2). The larger squared semi-partial correlation of the reading accuracy measure suggests this was the most important predictor. None of the other variables

explained unique variance in reading comprehension. Many variables previously found to predict reading comprehension, such as WM complex span (WMTB-C count recall) and inferencing ability, were not unique, significant predictors in this study. Possible explanations for the findings presented here will be discussed later.

Table 5 shows results of a standard multiple regression analysis with listening comprehension, as assessed using the Listening Comprehension Test Series, as the criterion variable.

Table 5: Standard regression analysis with LCTS listening comprehension as DV and language, memory and intelligence variables as predictors

Variable	B	SE B	Beta	t	p	sr ² (unique)
NARA-II accuracy	.031	.022	.115	1.394	.166	
Age	.029	.104	.017	.279	.781	
BPVS-II	.142	.042	.308**	3.397	.001	.036
TROG-2	.181	.144	.107	1.254	.212	
TOWK	.091	.085	.097	1.076	.284	
TRT	.217	.146	.104	1.484	.140	
Inferencing	.402	.157	.191*	2.567	.012	.020
Raven's CPM	.214	.099	.168*	2.165	.032	.014
WMTB-C digit recall	-.072	.110	-.046	-.649	.517	
WMTB-C count recall	-.032	.093	-.023	-.342	.733	

** $p < .01$, * $p < .05$

R for regression was significantly different to zero, $F(10, 114) = 20.919$, $p < .001$, with R^2 at .647 and adjusted R^2 at .616.

62% of the variance in listening comprehension was predicted by the predictor variables tested. Only three regression coefficients differed significantly from zero: vocabulary (BPVS-II), inferencing and non-verbal intelligence (Raven's CPM). The small squared semipartial correlations indicate that a large amount of variance in listening comprehension was shared by the predictors tested. None of the other

variables explained statistically significant unique variance in listening comprehension.

Summary of regression analyses

Table 6 summarises the results of the regression analyses conducted by highlighting those variables identified as unique, significant predictors of comprehension in each modality.

Table 6: Summary of standard regression analyses showing beta coefficients for variables predicting unique and significant variance in comprehension in each modality

Variable	NARA-II reading comprehension	LCTS listening comprehension
NARA-II accuracy	.652***	
Age		
BPVS-II	.167**	.308**
TROG-2	.123*	
TOWK		
TRT		
Inferencing		.191*
Raven's CPM		.168*
WMTB-C digit recall		
WMTB-C count recall		

*** $p < .001$, ** $p < .01$, * $p < .05$

Reading comprehension and listening comprehension appeared to be predicted by different variables. Whilst both depended on vocabulary knowledge, reading comprehension was also predicted by word recognition skills and syntactic skills, whilst listening comprehension was predicted by non-verbal intelligence and inferencing ability.

Regression analyses using residualised measures of comprehension

Rationale

The regression analyses presented above show the variables which explain unique variance in reading comprehension and listening comprehension. Of particular interest in this study, however, is the variance in comprehension in one modality which is not shared with variance in comprehension in the other modality. The Simple View of Reading predicts that reading comprehension can be explained by listening comprehension and word recognition ability only. Thus, having controlled for listening comprehension, it would be expected that reading comprehension would be predicted by word recognition ability only. Furthermore, the assumption that listening comprehension involves exactly the same processes as reading comprehension, apart from word recognition skills, suggests that, having controlled for reading comprehension, listening comprehension will not be systematically predicted by any of the predictor variables.

To examine variance in comprehension in one modality whilst controlling for comprehension in the other modality, residualised measures of reading comprehension and listening comprehension were created. The residualised measures of reading comprehension were an index of the amount of variance in reading comprehension not explained by listening comprehension, whilst the residualised measures of listening comprehension were an index of the amount of variance in listening comprehension not explained by reading comprehension. The analysis of residuals has previously been used in comprehension research in both adults (Long & Prat, 2008) and children (Nation & Snowling, 2004). For example, Nation and Snowling used the technique to examine the variance in word and exception word reading which could not be explained by non-word reading ability.

Regression was used to obtain residualised measures of reading comprehension and listening comprehension. Firstly, reading comprehension was regressed onto listening comprehension and the Pearson standardised residuals for each individual were saved to give a measure of reading comprehension independent of listening comprehension. For each individual, the reading comprehension residual reflected the standardised distance between an individual's actual reading comprehension score and the

regression line. Positive values occurred when reading comprehension was higher than the level predicted by listening comprehension, whilst negative values occurred when reading comprehension was lower than that predicted by listening comprehension. The larger the value of the residual, the greater the discrepancy between the actual value of reading comprehension obtained and that predicted from listening comprehension.

Similarly, listening comprehension was regressed onto reading comprehension and the Pearson standardised residuals saved to give a measure, for each individual, of listening comprehension independent of reading comprehension. Again, these values reflected, for each individual child, the discrepancy between the actual score for listening comprehension obtained and the score predicted by reading comprehension ability. Positive values occurred when listening comprehension was higher than predicted from reading comprehension, negative values when the reverse was true.

By obtaining residualised measures of reading comprehension and listening comprehension in this way, it was possible to explore the variance in comprehension in each modality not shared with comprehension in the other modality. To ascertain which, if any, of the variables explained unique variance in the residualised measures of comprehension, two standard multiple regression analyses were conducted with residualised reading comprehension (independent of listening comprehension) and residualised listening comprehension (independent of reading comprehension) as the criterion variables.

Results

Table 7 shows the results of a standard multiple regression analysis with reading comprehension residuals as the criterion variable.

Table 7: Standard regression analysis with NARA-II reading comprehension residuals as DV and language, memory and intelligence variables as predictors

Variable	B	SE B	Beta	t	p	sr ² (unique)
NARA-II accuracy	.034	.004	.771***	8.401	<.001	.269
Age	-.021	.019	-.075	-1.093	.277	
BPVS-II	-.004	.008	-.050	-.496	.621	
TROG-2	.020	.027	.069	.733	.465	
TOWK	-.001	.016	-.006	-.065	.948	
TRT	-.023	.027	-.068	-.869	.387	
Inferencing	-.081	.029	-.232**	-2.806	.006	.030
Raven's CPM	-.017	.018	-.082	-.946	.346	
WMTB-C digit recall	.021	.020	.082	1.041	.300	
WMTB-C count recall	.016	.017	.068	.918	.361	

*** $p < .001$, ** $p < .01$, * $p < .05$

R for regression was significantly different to zero, $F(10, 114) = 14.846$, $p < .001$ with R^2 at .566 and adjusted R^2 at .528.

53% of the variance in reading comprehension residuals could be explained by the predictor variables tested. Only two regression coefficients differed significantly from zero, word recognition (NARA-II accuracy) and inferencing. The positive beta value for word recognition suggested that, after controlling for all other variables, children whose reading comprehension was higher than that predicted by listening comprehension tended to have good word recognition ability, whilst those whose reading comprehension was lower than that predicted by listening comprehension tended to have poor word recognition ability. On the other hand, the negative beta value for inferencing suggested that, after controlling for all other variables, children whose reading comprehension was lower than that predicted by listening comprehension tended to have good inferencing skills, whilst those with reading comprehension higher than that predicted by listening comprehension tended to have poor inferencing skills. These findings deserve explanation and are explored further

later. None of the other variables explained unique variance in reading comprehension residuals.

Table 8 shows the results of a standard multiple regression analysis with listening comprehension residuals as the criterion variable.

Table 8: Standard regression analysis with LCTS listening comprehension residuals as DV and language, memory and intelligence variables as predictors

Variable	B	SE B	Beta	t	p	sr ² (unique)
NARA-II accuracy	-.019	.005	-.426***	-3.823	<.001	.082
Age	.018	.023	.063	.752	.453	
BPVS-II	.020	.009	.263*	2.152	.034	.026
TROG-2	.010	.032	.034	.294	.770	
TOWK	.012	.019	.076	.632	.529	
TRT	.043	.033	.123	1.297	.197	
Inferencing	.104	.035	.297**	2.956	.004	.049
Raven's CPM	.038	.022	.180	1.716	.089	
WMTB-C digit recall	-.023	.025	-.089	-.929	.355	
WMTB-C count recall	-.014	.021	-.062	-.691	.491	

*** $p < .001$, ** $p < .01$, * $p < .05$

R for regression was significantly different to zero, $F(10, 114) = 6.357$, $p < .001$ with R^2 at .358 and adjusted R^2 at .302.

30% of the variance in listening comprehension residuals could be explained by the predictor variables tested. Only three regression coefficients differed significantly from zero: vocabulary (BPVS-II), inferencing and word recognition (NARA-II accuracy). The positive beta values for vocabulary and inferencing showed that, after controlling for all other variables, children whose listening comprehension was higher than that predicted by reading comprehension tended to have good vocabulary and inferencing skills, whilst those whose listening comprehension was lower than that predicted by reading comprehension tended to have poor vocabulary and inferencing

skills. On the other hand, the negative beta value for word recognition suggested that, after controlling for all other variables, children whose listening comprehension was lower than that predicted by reading comprehension had good word recognition skills, whilst those with listening comprehension higher than that predicted by reading comprehension had poor word recognition skills. These findings are explored later. None of the other variables explained unique variance in listening comprehension residuals.

Summary of regression analyses using residualised measures of comprehension

Table 9 summarises the results of the standard regression analyses involving the residualised measures of comprehension in each modality by highlighting the unique, significant predictors of each residualised measure.

Table 9: Summary of standard regression analyses showing beta coefficients for variables predicting unique and significant variance in residualised measures of comprehension in each modality

Variable	NARA-II reading comprehension residuals, independent of LCTS listening comprehension	LCTS listening comprehension residuals, independent of NARA-II reading comprehension
NARA-II accuracy	.771***	-.426***
Age		
BPVS-II		.263*
TROG-2		
TOWK		
TRT		
Inferencing	-.232**	.297**
Raven's CPM		
WMTB-C digit recall		
WMTB-C count recall		

*** $p < .001$, ** $p < .01$, * $p < .05$

It can be seen that, having removed variance shared with listening comprehension, the only unique positive predictor of reading comprehension residuals was word recognition ability. Inferencing ability was a negative predictor of this residualised measure. Conversely, having removed variance shared with reading comprehension, additional variance in listening comprehension residuals was positively predicted by inferencing skill, but negatively predicted by word recognition ability. Vocabulary was a further unique positive predictor of residualised listening comprehension.

The presence of the negative predictors requires further explanation. Of note is the “symmetry” to the findings. Word recognition is a positive predictor of reading comprehension residuals but a negative predictor of listening comprehension residuals, whilst inferencing is a positive predictor of listening comprehension residuals and a negative predictor of reading comprehension residuals. To understand this phenomenon, the negative correlation between reading comprehension residuals and listening comprehension residuals needs to be taken into account ($r = -.665, p < .001$). This means that children whose reading comprehension scores are lower than predicted from their listening comprehension scores will tend to have listening comprehension scores that are higher than those predicted from their reading comprehension scores and vice versa. A variable that is a strongly positive predictor of comprehension residuals in one modality may, therefore, be a negative predictor of comprehension residuals in the other modality. For example, word recognition is a positive predictor of reading comprehension residuals. This means that high levels of this skill are shown by children achieving more highly on the reading comprehension test than predicted from their listening comprehension scores. The same children will tend to achieve less well on the listening comprehension test than predicted from their reading comprehension scores. Thus, high levels of word recognition skill are associated with highly positive reading comprehension residuals and highly negative listening comprehension residuals. On the other hand, low levels of word recognition skill are shown by children achieving less highly on the reading comprehension test than predicted from listening comprehension. These children will tend to achieve more highly on the listening comprehension test than predicted from their reading comprehension scores. Thus, low levels of word recognition skill are associated with highly negative reading comprehension residuals and highly positive listening comprehension residuals.

Figures 2 and 3 illustrate this explanation. In both graphs, the residualised measures of reading comprehension are plotted against the residualised measures of listening comprehension. The overall negative relationship between these two variables can be seen clearly. High scores on reading comprehension residuals tend to be accompanied by low scores on listening comprehension residuals and vice versa. Thus, as mentioned above, children who achieve more highly on the reading comprehension test than predicted from listening comprehension scores will tend to score lower on the listening comprehension test than predicted from reading comprehension scores. In Figure 2 a median split has been used to divide the children into those with good and poor word recognition skills, whilst in Figure 3 a median split divides the children into those with good and poor inferencing skill.

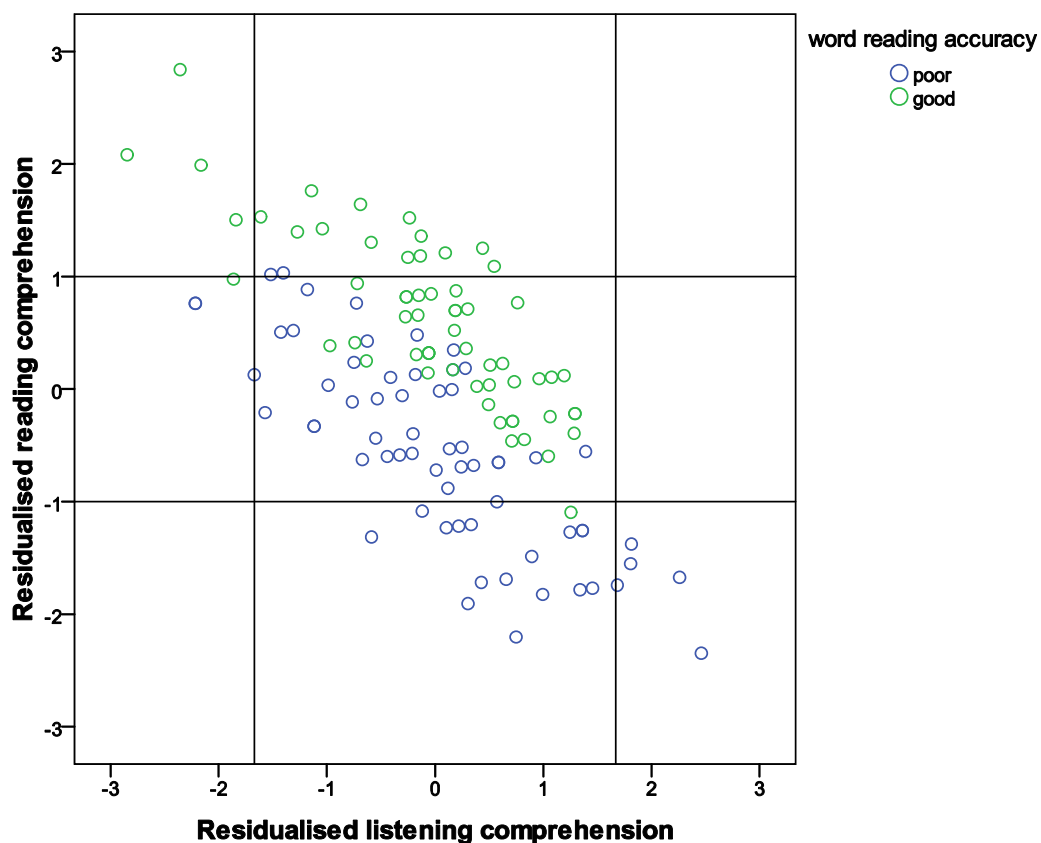


Figure 2: Relationship between residualised measures of reading comprehension and listening comprehension, with distribution of children with good and poor word recognition skills shown

As can be seen in Figure 2, children whose reading comprehension is better than that predicted from listening comprehension scores (standardised reading comprehension residuals > 1) tend to have good word recognition skills, whilst those whose reading comprehension is worse than that predicted from listening comprehension scores (standardised reading comprehension residuals < -1) tend to have poor word recognition skills. Furthermore, all children whose listening comprehension is considerably better than that predicted from reading comprehension scores (standardised listening comprehension residuals > 1.67) have poor word recognition skills whilst those whose listening comprehension is considerably worse than that predicted from reading comprehension scores (standardised listening comprehension residuals < -1.67) tend to have good word recognition skills.

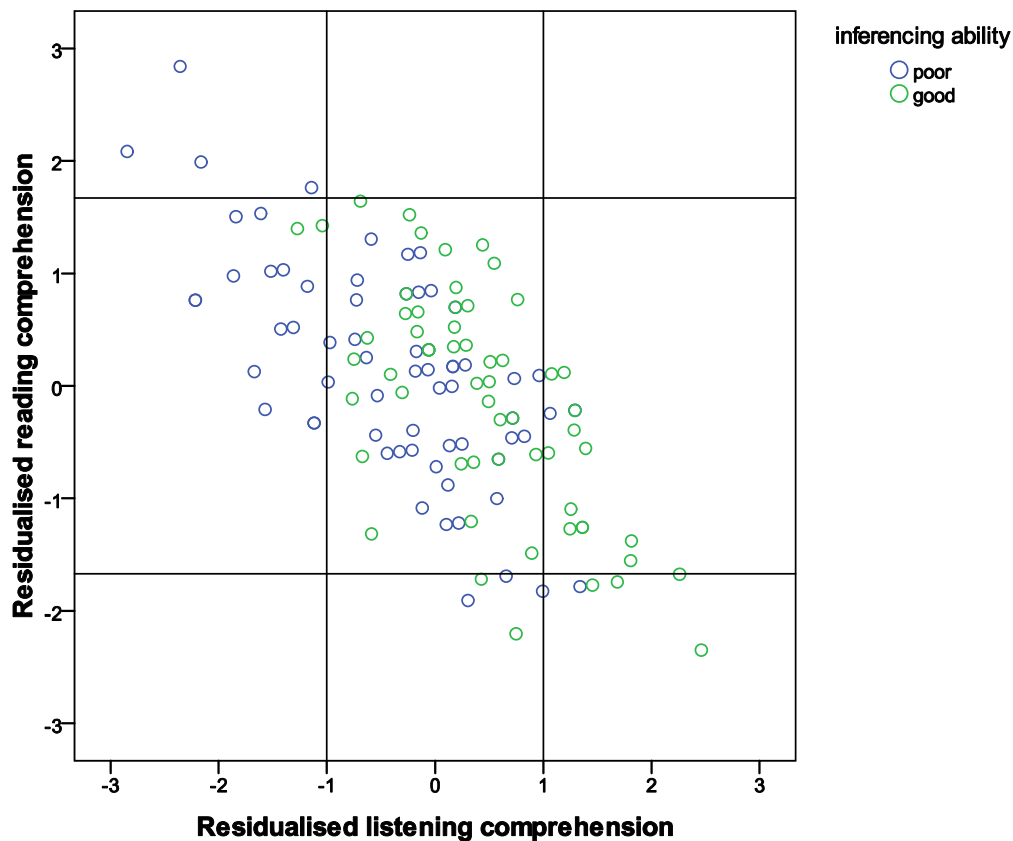


Figure 3: Relationship between residualised measures of reading comprehension and listening comprehension, with distribution of children with good and poor inferencing skills shown

Figure 3 clearly shows that children who are achieving more highly on listening comprehension than predicted from reading comprehension scores (standardised listening comprehension residuals > 1) tend to have good inferencing skills whilst those whose listening comprehension is worse than that predicted from reading comprehension scores (standardised listening comprehension residuals < -1) tend to have poor inferencing skills. Furthermore, all children whose reading comprehension is considerably better than that predicted from listening comprehension scores (standardised reading comprehension residuals > 1.67) have poor inferencing skills whilst those whose reading comprehension is considerably worse than that predicted from listening comprehension scores (standardised reading comprehension residuals < -1.67) tend to have good inferencing skills.

In summary, the presence of the negative predictors reflects the much greater importance of word recognition in reading comprehension than in listening comprehension and the much greater importance of inferencing ability in listening comprehension than in reading comprehension. Since the residualised measures of comprehension have a strong, negative correlation with each other, the presence of negative predictors becomes inevitable.

4.4. Discussion

The first aim of the study was to identify and compare those skills which made independent contributions to reading comprehension and those making independent contributions to listening comprehension. Standard regression analyses with reading comprehension and listening comprehension as criterion variables were conducted to fulfil this aim and the results are discussed in the first part of this Discussion. The second aim of the study was to identify which cognitive-linguistic skills continued to predict unique and significant variance in comprehension in each modality after controlling for comprehension in the other modality. Standard regression analyses using residualised measures of reading comprehension and listening comprehension, independent of comprehension in the other modality, were conducted. Results of these analyses are discussed in the second part of this Discussion.

1. Which skills account for unique variance in reading comprehension and which account for unique variance in listening comprehension?

The first set of regression analyses were conducted to find the most important predictors of reading comprehension and of listening comprehension. Both reading comprehension and listening comprehension were found to be uniquely predicted by vocabulary. However, the remaining variables predicting unique variance in comprehension differed for the two modalities. Reading comprehension was predicted by the word- and sentence-level language skills of word recognition accuracy and syntactic abilities, whilst listening comprehension was predicted by the higher cognitive skills of non-verbal intelligence and inferencing ability. The different pattern of findings for listening comprehension and reading comprehension suggests that, when using standardised tests, listening comprehension and reading comprehension do not make exactly the same demands on comprehenders. Apart from vocabulary, different abilities account for unique variance in comprehension in the two modalities. If, as assumed by the Simple View of Reading, exactly the same cognitive-linguistic processes underlie reading comprehension and listening comprehension, it would be expected that comprehension in each modality would be uniquely explained by the same predictor variables with additional variance in reading comprehension being explained by word recognition.

It was interesting to find that many of the variables identified in the literature as predictors of reading comprehension did not make unique, significant contributions to reading comprehension in this study. For example, neither working memory (complex span) nor inferencing ability appeared to be unique predictors of reading comprehension. Yet, as described in Chapter 3, the relationship between working memory and reading comprehension ability has been widely reported in the comprehension literature (e.g. Seigneuric, Ehrlich, Oakhill, & Yuill, 2000; Oakhill, Cain, & Bryant, 2003), as has the relationship between inferencing and reading comprehension (Oakhill, 1984; Oakhill, Cain, & Bryant, 2003). There are several possible explanations for the inability of these variables to explain variance in reading comprehension in the present study. Firstly, little previous research has considered potential predictor variables whilst statistically controlling for so many other cognitive-linguistic abilities. For example, to the author's knowledge, inferencing ability and syntactic skills have not previously been entered simultaneously into a

regression analysis. In line with findings from the current study, when many cognitive-linguistic variables have been considered simultaneously, complex working memory span has not always been found to be a predictor of reading comprehension. For example, Goff, Pratt, and Ong (2005) found that complex working memory span did not account for variance in reading comprehension in a normal sample of children having controlled for reading accuracy, vocabulary and syntactic skills. A further explanation for the findings of the current study arises from a consideration of the nature of the measures of inferencing and working memory used. In the current study, a numerical measure of working memory span was chosen. Findings of a relationship between numerical measures of working memory span and reading comprehension are much less robust than the findings associating linguistic measures of working memory span and reading comprehension (Cain, 2006; Leather & Henry, 1994). Indeed, in the study mentioned above, Goff et al. (2005) used a numerical measure of complex working memory span. Furthermore, it should be noted that in the current study, the test of inferencing ability was administered aurally whereas, in most previous studies, inferencing ability has been assessed using text which the children read themselves (Oakhill, 1984). A final factor which may be impacting on the results of the current study is the sampling method used. As in the study carried out by Goff et al. (2005), a relatively unselected sample of children was used. The high dependence of reading comprehension on reading accuracy found in the current study reflected this sampling method. In previous studies, however, poor readers have been excluded from the analysis (Oakhill, Cain, & Bryant, 2003) and the dependence of reading comprehension on reading accuracy has been found to be much lower.

2. Which skills account for unique variance in comprehension in each modality after controlling for comprehension in the other modality?

The second set of regression analyses, using residualised measures of reading comprehension and listening comprehension, were conducted specifically to investigate the variance in comprehension in one modality which was not shared with variance in comprehension in the other modality. As suggested by the Simple View of Reading, reading comprehension was positively predicted by word recognition ability only, once listening comprehension was controlled for. Systematic variance in NARA-II reading comprehension can be explained by word recognition ability and performance in the Listening Comprehension Test Series. This is a particularly strong

finding given the fact that the materials used to assess reading comprehension and listening comprehension were very different, a point that will be returned to later.

Listening comprehension continued to be positively predicted by vocabulary and inferencing skill once variance shared with reading comprehension was removed. The Simple View of Reading does not account for the systematic prediction of variance in listening comprehension by any other cognitive-linguistic skills after controlling for reading comprehension. The assumption that exactly the same processes are involved in reading comprehension and listening comprehension is not supported in this study as it would appear that, just as NARA-II reading comprehension makes extra demands on comprehenders in terms of their word recognition ability, so performance on the Listening Comprehension Test Series makes extra demands on listeners' vocabulary and inferencing skills above and beyond those shared with NARA-II reading comprehension.

Various methodological explanations for these findings could be proposed. One possible explanation is that the Listening Comprehension Test Series simply includes more difficult vocabulary and more questions requiring inferencing than the NARA-II. As far as vocabulary is concerned, it should be noted that the Listening Comprehension Test Series is standardised for children of this age whilst the NARA-II is standardised for children aged 6-12. The readers with good word recognition skills are, therefore, exposed to passages with vocabulary appropriate for children much older than themselves. The children with poor word recognition skills, on the other hand, will not have chance to read these more difficult passages. As far as inferencing is concerned, it seems unlikely that the Listening Comprehension Test Series includes more questions which require inferences to be made than the NARA-II. Analysis of the NARA-II questions themselves has demonstrated that performance on the test is heavily reliant on the generation of knowledge-based inferences (Bowyer-Crane & Snowling, 2005). Furthermore, children who are identified as poor comprehenders on the NARA-II have been shown to have difficulty on the questions involving knowledge-based inferencing (Bowyer-Crane & Snowling, 2005), and a wealth of research suggests that they also have problems on other tests of knowledge-based inferencing (Cain & Oakhill, 1999; Cain, Oakhill, & Bryant, 2004; Oakhill, 1984). The findings of the current study do not dispute these previous results. The

bivariate correlation between NARA-II comprehension and inferencing ability suggests that the better comprehenders do perform better on the aurally administered test of inferencing. What the findings of the present study suggest, however, is that the relationship between reading comprehension and the test of inferencing ability is mediated entirely by other variables.

A second explanation for the finding that performance on the Listening Comprehension Test Series appears to make greater demands on children's inferencing ability than the NARA-II, concerns the shared demands of the aurally-administered inferencing task and the Listening Comprehension Test Series. When assessed on the NARA-II, children were able to look back at the text to find answers to comprehension questions. In both the inferencing test and the Listening Comprehension Test Series, children could not refer to the text to answer questions. Therefore, the relationship between performance on the Listening Comprehension Test Series and the ability to answer questions requiring inference generation may simply have been due to the shared memory demands of the two tasks.

The results of this study suggest that, as proposed by the Simple View of Reading, performance on a standardised test of reading comprehension, the NARA-II, can be explained by children's word recognition skills and their performance on a standardised test of listening comprehension, the Listening Comprehension Test Series. The Listening Comprehension Test Series, on the other hand, does not appear to assess only those aspects of comprehension which are shared with the NARA-II. However, the fact that these two standardised tests make very different demands on comprehenders means that it is difficult to interpret the findings. For example, it is hard to know whether listening comprehension actually makes additional demands on children's inferencing skills and vocabulary knowledge compared to reading comprehension or whether the findings presented here simply reflect differences in the materials used. Similarly, it is hard to know whether, compared to the NARA-II, the Listening Comprehension Test Series makes extra demands on children's inferencing skills themselves or simply makes extra demands on the ability to remember information. The study reported in the following chapter attempted to answer these questions by using non-standardised tests of reading comprehension and

listening comprehension which used similar materials and made similar demands on children's memory.

Chapter 5.

Study 2: Are listening comprehension and reading comprehension underpinned by the same cognitive-linguistic skills? An investigation using true/false tests.

5.1. Introduction

Results from Study 1 suggested that reading comprehension and listening comprehension are predicted by different variables, and that listening comprehension makes demands on comprehenders that are additional to those made by reading comprehension. However, as discussed previously, a problem in the interpretation of the results arises from the fact that the standardised tests used in Study 1 differ in several important ways.

Firstly, the materials themselves are different in the two standardised tests. Whilst the NARA-II is comprised of narrative passages only, the Listening Comprehension Test Series assesses listening comprehension using a range of different genres, including a poem and a conversation. It has been argued that in order to compare reading comprehension and listening comprehension, tests should be equivalent apart from the input modality (Hedrick & Cunningham, 1995; Hoover & Gough, 1990). Indeed, many of the studies considering the relationship between reading comprehension and listening comprehension that were reviewed in Chapter 2 have taken this approach (e.g. Curtis, 1980; Hoover & Gough, 1990; Megherbi, Seigneure, & Ehrlich, 2006).

Secondly, when assessed on the NARA-II, children are allowed to look back at the text to answer questions, whilst it is obviously impossible for them to re-visit the text whilst answering questions when assessed on the Listening Comprehension Test Series. It should be noted, however, that the Listening Comprehension Test Series is administered in such a way that children hear a passage twice before having to attempt the questions, and hear the questions themselves before their second exposure to the passage. According to the test manual, giving the listeners an awareness of “what to listen for” closely replicates real-life listening situations as listening tends to occur in a context (Hagues, Siddiqui, & Merwood, 1999). Nevertheless, on balance, it

might be expected that demands on memory for the text are higher in the standardised test of listening comprehension than in the standardised test of reading comprehension. As explained previously, this could explain the greater dependence of listening comprehension on performance in an aurally administered test of inferencing.

Study 2, therefore, aimed to compare reading comprehension and listening comprehension using tests of comprehension that made demands that were more comparable than those made by the Listening Comprehension Test Series and the NARA-II. Specifically, it was important that the tests used similar materials and made similar memory demands on comprehenders.

One approach that has sometimes been taken in previous UK studies of comprehension is the administration of the Neale Analysis of Reading Ability as a test of both reading comprehension and listening comprehension (Cain, Oakhill, & Bryant, 2000a; Stothard & Hulme, 1992). Two parallel standardised versions of the test are available: Form 1 and Form 2. In previous studies, one version has been administered as a reading comprehension test, whilst the other version has been administered as a listening comprehension test in which the stories and questions are read aloud to the children. The use of the standardised test of reading comprehension in both modalities means that the materials involved in assessing reading comprehension and listening comprehension are very similar.

A different approach has been taken by Spooner, Baddeley, and Gathercole (2004) who raised concerns about aspects of administration of the NARA-II. In particular, they argued that children's comprehension ability may be underestimated by the open-ended nature of the questions asked in the test. They identified two groups, matched for word reading ability but differing on reading comprehension scores, by administering Form 1 of the NARA-II according to standardised instructions. They found that these two groups did not differ on comprehension scores when Form 2 of the NARA-II was administered in such a way that the open-ended questions were replaced with forced choice true/false questions. Spooner et al. suggested that children identified by the NARA-II as poor comprehenders may in fact simply struggle with the expressive language demands made by open-ended questions. To overcome this

issue, Spooner et al. designed a test to assess both reading comprehension and listening comprehension which requires forced-choice true/false responses to avoid confounds with expressive language skills. The test uses similar material to assess comprehension in each modality and makes the same demands on memory whether administered as a reading comprehension or listening comprehension test. This test was chosen for use in Study 2.

Two parallel versions of the test have been developed, both of which can be used to assess comprehension in either modality. Each version of the test consists of six stories of increasing length and difficulty. The materials involved are largely adapted from the NARA-II itself. The NARA-II includes a Diagnostic Tutor Form which provides additional comprehension stories and questions. In their true/false tests, Spooner et al. have used stories from the Diagnostic Tutor Form and from Form 2 of the NARA-II, and have also included a small number of stories especially written for the test. All questions assessing comprehension require forced-choice true/false responses. Where stories have been taken from the NARA-II, an attempt has been made to retain the content of the original questions and their answers. As in the original test, some questions are designed to assess the ability to infer information from the text whilst others assess the ability to retrieve explicitly stated information. To ensure that demands on memory are similar in each modality, when reading comprehension is assessed children must answer the questions without referring to the text that they have read. It has been argued that the use of tests of reading comprehension in which the text is not available at questioning reduces the ecological validity of a study (Goff, Pratt, & Ong, 2005). In this study, however, it was important to ensure that the reading comprehension test made comparable memory demands to the listening comprehension test.

The test developed by Spooner et al. has not been standardised. Furthermore, it has been criticised on several grounds (Cain & Oakhill, 2006b). Cain and Oakhill argue that true/false questions are limited in their ability to assess whether inferences have actually been generated by the comprehender, rather than simply recognised at test. Furthermore, they express concern that the test lacks sensitivity with scores reported by Spooner et al. approaching ceiling. Spooner et al. administered only the first 4 stories of their test, a total of 28 items. In the current study, all 6 stories were

administered, a total of 44 items, to avoid ceiling effects and improve task sensitivity. Because this study involved the same children who had already completed standardised tests of comprehension, it afforded an opportunity not only to explore the relationship between listening comprehension and reading comprehension when task demands were comparable but also to explore the relationship of the non-standardised, forced-choice true/false measures to the standardised measures of comprehension.

5.2. Summary of aims

The first aim of this study was to identify and compare skills making independent contributions to reading comprehension and to listening comprehension when comprehension was assessed using true/false tests. The tests had been designed to allow comprehension in each modality to be compared when materials and task demands were similar. It was predicted that, as suggested by the Simple View of Reading, comprehension in each modality would be uniquely explained by the same predictor variables with additional variance in reading comprehension being explained by word recognition ability. The second aim was to identify those skills making independent contributions to true/false comprehension in each modality after controlling for true/false comprehension in the other modality. It was predicted that, as suggested by the Simple View of Reading, reading comprehension would be explained by word recognition ability only, once listening comprehension had been controlled for, and that listening comprehension would not be uniquely explained by any cognitive-linguistic skills, having controlled for reading comprehension. The third aim of the study was to explore the relationship between the true/false tests of comprehension and the standardised tests, used in Study 1, by comparing patterns of results obtained using both types of test.

5.3. Method

Design

Two standard regression analyses were conducted with true/false reading comprehension and true/false listening comprehension as criterion variables. Predictor variables entered into each regression were word recognition accuracy (NARA-II), age, vocabulary (BPVS-II), syntactic ability (TROG-2), semantic awareness

(TOWK), exposure to print (TRT), inferencing ability, non-verbal intelligence (Raven's CPM), short-term memory (WMTB-C digit recall) and working memory (WMTB-C count recall). True/false comprehension scores in each modality were then regressed onto true/false comprehension scores in the other modality and the residuals saved. These residualised measures of comprehension were entered as criterion variables in two further standard regression analyses. Predictor variables were those listed above.

Participants

The sample of children taking part in Study 1 also took part in Study 2.

Materials

Reading Comprehension and Listening Comprehension

Forced-choice true/false tests of listening comprehension and reading comprehension, developed by Spooner, Baddeley, and Gathercole (2004) were used.

In these tests children listen to or read six passages of increasing length and difficulty. In the reading test, the passages are presented in a booklet. Children read each passage silently and without assistance. For each passage, an illustrative picture appears on the facing page. After each passage, children turn the page and read a series of statements relating to the text. They have to indicate whether each statement is true or false by placing a tick or a cross in a box at the end of each one. They are not allowed to refer back to the text whilst doing this. The illustrative picture also appears on the page facing the true/false questions. Whilst providing a broad context for the story, the picture does not provide information that can be used to answer any of the questions. In the listening test, children hear each passage, read slowly and clearly by the experimenter. After each passage, they hear a series of statements and again have to decide whether each is true or false. Answers are recorded by placing a tick or a cross in an appropriately numbered response box. Again, children are given booklets in which to record their answers. For each story, the numbered response boxes are presented on a page which faces an illustrative picture.

For both reading and listening versions of the task, four questions accompany the shortest story, whilst eight questions are asked about the subsequent ones. Children

are encouraged to attempt all questions even if this means that they have to guess the answers.

An example of a story used in the true/false test of listening comprehension is shown in Table 10. All stories and questions used in the true/false tests of listening comprehension and reading comprehension are presented in Appendices 4 and 5.

Table 10: Example story from true/false test of listening comprehension

Tony and Susan awoke suddenly. The dog was barking loudly in the yard. The children ran to the window. They could see smoke and flashing lights some way off. A helicopter had crash-landed in the park nearby. Flames shot into the air. They saw the pilot jump clear and run to safety.

True or false?

Tony and Susan were woken up by the dog running into their room.	false
They ran to the yard.	false
They saw lights and smoke.	true
A helicopter had exploded.	false
The crash happened in the park.	true
The helicopter was on fire.	true
The pilot was in danger because he was badly hurt.	false
The pilot jumped out of the helicopter.	true

Predictor Variables

All predictor variables were assessed using the measures used in Study 1 and described in Chapter 4.

Procedure

The true/false tests of listening comprehension and reading comprehension were administered to whole classes over two half-hour sessions which were held about a week apart.

In the first session, the true/false test of listening comprehension was administered. A practice story was first read aloud to the class by the experimenter followed by four true/false statements which the class worked through together. The true/false test of reading comprehension was administered in the second session, at which point children were familiar with the format of the questions and could work independently through their booklets.

5.4. Results

Scoring

In previous studies in which these true/false tests have been used (Spooner et al., 2004) performance has been scored as the number of items correctly identified as true or false. It has been argued, however, that a problem with scoring forced-choice true/false tests in this way is that it confounds two separate factors, sensitivity and response bias (Stanislaw & Todorov, 1999). In the case of the comprehension tests under consideration, sensitivity is a measure of how well a child actually discriminates between statements that are true and those that are false. This is the measure that is of interest. Response bias, on the other hand, is the general tendency of a child to respond “true” or “false”. Royer, Hastings, and Hook (1979) argue that, if the total number of correct responses is taken as a measure of performance in two-choice discrimination problems such as this, it is not clear whether two children with different scores differ in actual sensitivity, response bias or both. It may be that one child is willing to respond “true” when they have only a slight feeling that the idea expressed in the statement is one that appeared previously, whilst another child may need to feel much more confident before responding “true”. A situation could arise in which the two children’s sensitivity to the items is identical but their performance very different because of their different response biases.

A measure of sensitivity can, however, be calculated which is independent of response bias. This is the d-prime score (d'). The computation of d-prime scores is appropriate whenever participants have to discriminate between two types of stimulus, one of which is seen as a “signal” whilst the other is seen as “noise” (Stanislaw & Todorov, 1999). For example, Royer, Hastings, and Hook (1979) argue that

calculation of d-prime scores is appropriate when using the Sentence Verification Technique. This a measure of comprehension in which children have to respond “old” or “new” to sentences according to whether or not they have the same meaning as sentences which have previously been presented in a short passage. In this case, “old” statements can be regarded as “signal” stimuli whilst “new” statements can be regarded as “noise”. Furthermore, in their study Royer et al. demonstrated that the d-prime measure led to a more “powerful” analysis than a simple proportion correct score, in that it allowed deviation from the null hypothesis to be detected more readily.

In the current study, involving forced-choice, true/false comprehension tests, “true” statements were regarded as “signal” stimuli whilst “false” statements were regarded as “noise”. D-prime scores were obtained for each child on both the true/false listening comprehension test and the true/false reading comprehension test. The d-prime scores were calculated using the formula $d' = z(\text{Hits}) - z(\text{False Alarms})$, where “Hits” was defined as the proportion of “true” responses given to “true” items and “False Alarms” was defined as the proportion of “true” responses given to “false” items. The higher the d-prime score, the greater the child’s ability to discriminate between true and false items. Thus, for each child, scores measuring their actual sensitivity on each of the true/false tests of comprehension were obtained, guessing being corrected for. All subsequent analyses were conducted using these d-prime scores.

Sample size

As demonstrated in Chapter 4, a sample size of 126 was adequate for testing the multiple correlation and the individual predictors. There was a small amount of missing data (0.07%) which was dealt with by excluding cases listwise from the analysis. Because the distribution of the missing data was slightly different in Studies 1 and 2, slightly different samples were used in the analyses reported in the two studies.

Data preparation

For reasons explained in Chapter 4, it is important that data is rigorously screened when regression techniques are used.

The distributions of the d-prime scores for the true/false measures of comprehension were explored for univariate outliers. On the true/false test of listening comprehension, the d-prime scores for two of the cases were slightly greater than 3.29 standard deviations above the mean d-prime score. As in Study 1, these outlying scores were retained for analysis as there was no evidence that any case was exerting an undue influence on the results (see below). Values for skew and kurtosis were calculated. Whilst on the true/false test of reading comprehension, these values were found to be between +1 and -1, kurtosis for the d-prime scores on the true/false test of listening comprehension was found to be $> +1$. This was not regarded as a problem, however, as positive kurtosis does not lead to an underestimate of the variance when samples include more than 100 cases (Tabachnick & Fidell, 2007).

For each of the regression analyses reported in this chapter, residuals were examined. Normal probability plots and scatterplots of predicted values of the DV against standardised residuals showed no major deviations from normality, linearity and homoscedasticity. Furthermore, Cook's distances obtained showed that no case appeared to exert undue influence on the results. However, in some of the regression analyses conducted, an outlier was found in the solution. This is discussed further when the analyses are reported.

Descriptive statistics

Means and standard deviations for d-prime scores on the true/false tests of comprehension in each modality were obtained and are shown in Table 11.

Table 11: Means and standard deviations for d-prime scores on forced-choice measures of comprehension

Variable	N	Mean	SD	Skew	Kurtosis
True/false reading comprehension (d' max = 4.65, min = 0.00)	125	1.24	0.70	-.018	-.714
True/false listening comprehension (d' max = 4.65, min = 0.00)	126	1.67	0.59	.448	1.477

Reliability

For the true/false test of reading comprehension, split half reliability of the d-prime scores (Spearman-Brown correction) was calculated as .569. For the true/false test of listening comprehension, split half reliability of the d-prime scores (Spearman-Brown correction) was calculated as .505. The low reliability of these tests was a cause for concern and will be discussed later.

Correlations between comprehension measures

To explore relationships between the standardised tests of comprehension used in Study 1 and the true/false tests used in Study 2, bivariate correlations between the four measures of comprehension were obtained and are shown in Table 12.

Table 12: Bivariate correlations between measures of comprehension

	1	2	3	4
1. NARA-II		.652**	.728**	.505**
2. Listening Comprehension Test Series			.614**	.577**
3. True/false reading comprehension (d')				.440**
4. True/false listening comprehension (d')				

**p < .001

High correlations were found between performance on the standardised and true/false tests of comprehension in each modality ($r = .728$ and $r = .577$ for tests of reading comprehension and listening comprehension respectively).

Following the procedure advocated by Raghunathan, Rosenthal, and Rubin (1996) for comparing two nonoverlapping correlations obtained from the same sample, the correlation between performance on the standardised tests of reading comprehension and listening comprehension ($r = .652$) was found to be significantly greater than the correlation between d-prime scores on the true/false tests of reading comprehension and listening comprehension ($r = .440$) ($Z_{PF} (N = 124) = 2.860, p < .01$). In other words, significantly more variance was shared between the NARA-II and the Listening Comprehension Test Series than between the true/false tests of reading comprehension and listening comprehension. This was surprising given that the task demands of the true/false tests of reading comprehension and listening comprehension were much more similar than those of the standardised tests of reading comprehension and listening comprehension.

Regression analyses

To ascertain which variables explained unique variance in the true/false tests of comprehension, two standard multiple regression analyses were conducted, with d-prime scores obtained on these tests as the criterion variables.

Because the sample was slightly different to that used in Study 1, bivariate correlations (two-tailed) between all variables were again obtained to check for evidence of multicollinearity between the predictor variables. Results are shown in Table 13. A Bonferroni correction was applied, giving a criterion for significance of $p < .00076 (.05/66)$.

Table 13: Correlations between Study 2 measures

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1.True/false LC d'	-	.440**	.449**	.149	.591**	.446**	.495**	.348**	.391**	.499**	.327*	.205
2.True/false RC d'		-	.688**	.180	.650**	.628**	.621**	.440**	.392**	.458**	.329*	.372**
3.NARA-II accuracy			-	.212	.573**	.502**	.630**	.550**	.214	.421**	.414**	.351**
4.Age				-	.304*	.194	.323*	.094	.056	.189	.111	.254
5.BPVS-II					-	.573**	.680**	.423**	.524**	.558**	.441**	.389**
6.TROG-2						-	.612**	.430**	.458**	.621**	.504**	.390**
7.TOWK							-	.471**	.439**	.505**	.375**	.392**
8.TRT								-	.275	.387**	.383**	.304*
9.Inferencing									-	.454**	.119	.351**
10.Raven's CPM										-	.421**	.437**
11.WMTB-C digit recall											-	.355**
12.WMTB-C count recall												-

**p < .0001, *p < .001

Key: True/false LC d'- d-prime scores on true/false test of listening comprehension; True/false RC d'- d-prime scores on true/false test of reading comprehension; NARA-II accuracy – Word recognition from the Neale Analysis of Reading Ability (Second Revised Edition); Age – chronological age in months; BPVS-II – British Picture Vocabulary Scales (Second Edition); TROG-2 – Test for Reception of Grammar (Version 2); TOWK – Test of Word Knowledge (Word Definitions sub-scale); TRT – Title Recognition Test; Inferencing – Correct inference responses; Raven's CPM – Raven's Coloured Progressive Matrices (1998 Edition); WMTB-C digit recall – Working Memory Test Battery for Children, simple span task; WMTB-C count recall – Working Memory Test Battery for Children, complex span task

As found in Study 1, there was no evidence of multicollinearity between the predictor variables as none of the correlation coefficients for relationships between predictor variables exceeded the value of .80.

Table 14 shows results of a standard multiple regression analysis with d-prime scores on the true/false test of reading comprehension as the criterion variable.

Table 14: Standard regression analysis with d-prime scores on true/false reading comprehension as DV and language, memory and intelligence variables as predictors

Variable	B	SE B	Beta	t	p	sr ² (unique)
NARA-II accuracy	.013	.003	.407***	4.908	<.001	.076
Age	-.012	.012	-.060	-.964	.337	
BPVS-II	.015	.005	.280**	3.066	.003	.030
TROG-2	.063	.017	.319***	3.716	<.001	.044
TOWK	.005	.010	.049	.542	.589	
TRT	-.002	.017	-.008	-.109	.913	
Inferencing	.004	.019	.016	.207	.836	
Raven's CPM	-.008	.012	-.056	-.706	.481	
WMTB-C digit recall	-.024	.013	-.134	-1.853	.066	
WMTB-C count recall	.010	.011	.060	.892	.374	

***p < .001, **p < .01, *p < .05

R for regression was significantly different to zero, $F(10,114) = 20.154$, $p < .001$ with R^2 at .639 and adjusted R^2 at .607.

The linear model accounted for a substantial amount of variance (61%) in d-prime scores on the true/false test of reading comprehension. Only three regression coefficients differed significantly from zero: word recognition (NARA-II accuracy), vocabulary (BPVS-II) and syntactic abilities (TROG-2). The larger squared semi-partial correlation of the reading accuracy measure suggests this was the most

important predictor. None of the other variables explained unique variance in the criterion variable.

Table 15 shows results of a standard multiple regression analysis with d-prime scores on the true/false test of listening comprehension as the criterion variable.

Table 15: Standard regression analysis with d-prime scores on true/false listening comprehension as DV and language, memory and intelligence variables as predictors

Variable	B	SE B	Beta	t	p	sr ² (unique)
NARA-II accuracy	.004	.003	.136	1.296	.197	
Age	-.002	.013	-.015	-.191	.849	
BPVS-II	.015	.005	.335**	2.916	.004	.042
TROG-2	.000	.018	-.005	-.050	.961	
TOWK	.007	.011	.071	.624	.534	
TRT	.004	.018	.020	.221	.826	
Inferencing	.020	.020	.096	1.017	.311	
Raven's CPM	.028	.012	.220*	2.232	.028	.025
WMTB-C digit recall	.007	.014	.044	.487	.627	
WMTB-C count recall	-.020	.012	-.143	-1.691	.094	

***p < .001, **p < .01, *p < .05

R for regression was significantly different to zero, $F(10, 115) = 8.678$, $p < .001$, with R^2 at .430 and adjusted R^2 at .381.

38% of the variance in d-prime scores in true/false listening comprehension was predicted by the predictor variables tested. Only two regression coefficients differed significantly from zero: vocabulary (BPVS-II), and non-verbal intelligence (Raven's CPM). The small squared semipartial correlations indicate the large amount of variance in listening comprehension, accounted for by the regression model, that was shared by the predictors tested. None of the other variables explained unique variance in d-prime scores on the true/false listening comprehension test.

Table 16 summarises the results of the regression analyses conducted by showing those variables identified as unique, significant predictors of d-prime scores on the true/false tests of comprehension. For the sake of comparison, Table 16 also shows the variables identified in Study 1 as explaining unique and significant variance on the standardised tests of comprehension in each modality.

Table 16: Summary of standard regression analyses showing beta coefficients for variables predicting unique and significant variance in standardised and in true/false measures of comprehension

Variable	True/false reading comprehension	True/false listening comprehension	NARA-II reading comprehension	LCTS listening comprehension
NARA-II accuracy	.407***		.652***	
Age				
BPVS-II	.280**	.335**	.167**	.308**
TROG-2	.319***		.123*	
TOWK				
TRT				
Inferencing				.191*
Raven's CPM		.220*		.168*
WMTB-C digit recall				
WMTB-C count recall				

***p < .001, **p < .01, *p < .05

When assessed using the true/false measures of comprehension, which used similar materials and made similar demands on memory, d-prime scores on reading comprehension and listening comprehension were predicted by different variables. Both depended on vocabulary knowledge, but reading comprehension was also uniquely predicted by word recognition and syntactic abilities whilst listening comprehension was also uniquely predicted by non-verbal intelligence. This pattern of results was similar to that obtained when standardised measures of comprehension were used. Performance on both the NARA-II and the true/false test of reading

comprehension was predicted by word recognition, vocabulary knowledge and syntactic skills, whilst performance on both the Listening Comprehension Test Series and the true/false test of listening comprehension was predicted by vocabulary and non-verbal intelligence. However, whilst additional variance in the Listening Comprehension Test Series was predicted by inferencing ability, this was not the case for the true/false test of listening comprehension.

Regression analyses using residualised measures of comprehension

Using the d-prime scores obtained on the true/false tests, residualised measures of reading comprehension and listening comprehension were obtained in order to explore the variance in comprehension in one modality which was not shared with variance in comprehension in the other modality.

Residualised measures of true/false reading comprehension were obtained by regressing d-prime scores for true/false reading comprehension onto d-prime scores for true/false listening comprehension and saving the Pearson standardised residuals. For each individual, this gave a measure of true/false reading comprehension which was independent of true/false listening comprehension. Effectively, this measure reflected the difference between an individual's actual d-prime score obtained on the true/false test of reading comprehension and that predicted by their d-prime score on the true/false test of listening comprehension.

D-prime scores for true/false listening comprehension were then regressed onto d-prime scores for true/false reading comprehension, and the Pearson standardised residuals were saved. For each individual, this measure reflected the difference between their actual d-prime score obtained on the true/false test of listening comprehension and that predicted by their d-prime score on the true/false test of reading comprehension. When true/false listening comprehension was regressed onto true/false reading comprehension, one outlier was identified in the solution. For this case, the d-prime score obtained on the true/false listening comprehension test was not well predicted by the d-prime score obtained on the true/false reading comprehension test. Investigation of the Cook's distance for this case, however, suggested that it was not exerting an undue influence on the model as a whole and it was retained in the analysis.

In order to ascertain which of the predictor variables, if any, explained unique variance in residualised measures of true/false comprehension in each modality, two standard multiple regression analyses were conducted.

Table 17 shows results of a standard multiple regression analysis with true/false reading comprehension residuals, independent of true/false listening comprehension, as the criterion variable.

Table 17: Standard regression analysis with true/false reading comprehension residuals as DV and language, memory and intelligence variables as predictors

Variable	B	SE B	Beta	t	p	sr ² (unique)
NARA-II accuracy	.018	.005	.392***	3.667	<.001	.071
Age	-.016	.023	-.058	-.721	.472	
BPVS-II	.012	.009	.152	1.294	.198	
TROG-2	.102	.031	.360**	3.255	.001	.056
TOWK	.003	.018	.017	.150	.881	
TRT	-.007	.032	-.019	-.210	.834	
Inferencing	-.011	.034	-.033	-.337	.737	
Raven's CPM	-.036	.021	-.171	-1.693	.093	
WMTB-C digit recall	-.044	.024	-.171	-1.844	.068	
WMTB-C count recall	.032	.020	.138	1.587	.115	

***p < .001, **p < .01, *p < .05

R for regression was significantly different to zero, $F(10, 114) = 7.663$, $p < .001$, with R^2 at .402 and adjusted R^2 at .350. One case was not well predicted by this solution, but the Cook's distance suggested that this was not unduly affecting the model as a whole.

35% of the variance in true/false reading comprehension residuals was explained by the predictor variables tested. Only two regression coefficients differed significantly

from zero, word recognition (NARA-II) and syntactic ability (TROG-2). The positive beta values for these variables suggested that children whose d-prime scores on the true/false reading comprehension test were higher than those predicted by their d-prime scores on the true/false listening comprehension test tended to have better word recognition and syntactic abilities than those children who performed less well on the reading comprehension test than predicted from their performance on the listening comprehension test.

Table 18 shows results of a standard multiple regression analysis with true/false listening comprehension residuals as the criterion variable.

Table 18: Standard regression analysis with true/false listening comprehension residuals as DV and language, memory and intelligence variables as predictors

Variable	B	SE B	Beta	t	p	sr ² (unique)
NARA-II accuracy	-.003	.006	-.058	-.470	.639	
Age	.002	.026	.009	.095	.925	
BPVS-II	.017	.011	.226	1.651	.101	
TROG-2	-.047	.036	-.166	-1.291	.199	
TOWK	.010	.021	.060	.445	.657	
TRT	.010	.037	.028	.261	.795	
Inferencing	.037	.040	.106	.940	.349	
Raven's CPM	.058	.025	.276*	2.341	.021	.039
WMTB-C digit recall	.030	.028	.116	1.074	.285	
WMTB-C count recall	-.044	.024	-.190	-1.881	.062	

***p < .001, **p < .01, *p < .05

R for regression was significantly different to zero, $F(10, 114) = 2.650$, $p < .01$, with R^2 at .189 and adjusted R^2 at .117. Again, one case was not well predicted by this solution but, again, the Cook's distance suggested this was not a problem.

12% of the variance in residualised true/false listening comprehension was explained by the predictor variables tested. Only one regression coefficient differed significantly

from zero, non-verbal intelligence (Raven's CPM). Children whose d-prime scores were better on the true/false listening comprehension test than expected from their performance on the reading comprehension test tended to score highly on non-verbal intelligence. Those whose performance on the listening comprehension test was worse than expected from their reading comprehension tended to score poorly on non-verbal intelligence.

Table 19 summarises the results of the standard regression analyses of the residualised measures of true/false comprehension by showing those variables identified as unique, significant predictors of residualised true/false comprehension. For the sake of comparison, Table 19 also shows the variables identified in Study 1 as explaining unique and significant variance in residualised measures of comprehension obtained using standardised tests.

Table 19: Summary of standard regression analyses showing beta coefficients for variables predicting unique and significant variance in residualised standardised and true/false measures of comprehension

Variable	Residualised true/false reading comprehension	Residualised true/false listening comprehension	Residualised NARA-II reading comprehension	Residualised LCTS listening comprehension
NARA-II accuracy	.392***		.771***	-.426***
Age				
BPVS-II				.263*
TROG-2	.360**			
TOWK				
TRT				
Inferencing			-.232**	.297**
Raven's CPM		.276*		
WMTB-C digit recall				
WMTB-C count recall				

***p < .001, **p < .01, *p < .05

It can be seen that, having removed variance shared with true/false listening comprehension, true/false reading comprehension continued to be predicted by word recognition and syntactic skills. Having removed variance shared with true/false reading comprehension, additional variance in true/false listening comprehension was explained by non-verbal intelligence only.

The pattern of results obtained is different for the standardised and true/false tests. Word recognition predicts residuals of reading comprehension whether standardised or true/false tests are used. However, residuals of reading comprehension are also predicted by syntactic ability when true/false tests of comprehension are used. Similarly, only non-verbal intelligence is a unique predictor of listening comprehension residuals when assessed using true/false measures. The residuals of a standardised measure of listening comprehension, however, are uniquely predicted by vocabulary and inferencing ability.

5.5. Discussion

The first aim of this study was to identify and compare those skills making independent contributions to reading comprehension and to listening comprehension when comprehension was assessed using true/false tests. These true/false tests had been designed specifically to allow comprehension in each modality to be compared when materials and task demands were similar. The second aim was to identify those skills making independent contributions to comprehension in each modality after controlling for comprehension in the other modality when comprehension was assessed using true/false tests. The third aim of the study was to explore the relationship between these true/false tests of comprehension and the standardised tests used in Study 1 by comparing patterns of results obtained using both types of test. Thus, in the following Discussion, results obtained using the true/false measures are explored, as well as their relationship to results obtained using the standardised tests. The section ends with a brief note of caution concerning the reliabilities of the true/false tests.

1. What skills account for unique variance in reading comprehension and which skills account for unique variance in listening comprehension when true/false tests are used? How do these results compare with those obtained using standardised tests of comprehension?

The first set of regression analyses showed that reading comprehension, assessed using a true/false test, was uniquely predicted by word recognition, vocabulary and syntactic abilities. These were the same variables that had been found to predict NARA-II reading comprehension in Study 1. True/false listening comprehension was found to be uniquely predicted by vocabulary and non-verbal intelligence. These variables had also been found to be unique predictors of performance on the Listening Comprehension Test Series. However, further unique variance in performance on the Listening Comprehension Test Series had been predicted by inferencing ability.

The finding that inferencing was not a unique predictor of true/false comprehension in either modality was interesting. If, as speculated in Study 1, the importance of inferencing in performance on the Listening Comprehension Test Series was due to the shared memory demands of the two tasks, it would be expected that inferencing would be a unique predictor of both true/false listening comprehension and true/false reading comprehension. Both these tests make demands on comprehenders' memory skills because in neither can the text be referred to at test. As far as reading comprehension is concerned, therefore, it appears that even when memory demands are shared by the reading comprehension test and the inferencing test, inferencing is not identified as a unique predictor of comprehension skill.

The finding that inferencing ability predicts performance on the Listening Comprehension Test Series but does not predict performance on the true/false test of listening comprehension is difficult to explain. In both comprehension tests, children make forced-choice responses and questions are asked which require inferences to be made. Whilst in the true/false test all questions are of the true/false type, the standardised test utilises some true/false questions and some multiple choice questions which require children to select the correct response out of four possible options. The use of true/false questions to assess comprehension has been criticised (Cain & Oakhill, 2006b), because, it is argued, children's ability to recognise inferences as correct is not the same as their ability to generate inferences spontaneously. Whilst

Cain and Oakhill also make this criticism of tests involving multiple choice questions, it may be that the more varied format of the questions in the Listening Comprehension Test Series is simply more sensitive to children's ability to generate inferences than a test that is comprised entirely of true/false questions.

2. Which skills account for unique variance in comprehension in each modality after controlling for comprehension in the other modality when using true/false comprehension tests? How do these results compare with those obtained using standardised tests of comprehension?

The second set of regression analyses showed that, when using true/false tests, reading comprehension continued to be predicted by word recognition and syntactic skills after controlling for listening comprehension. According to the Simple View of Reading, all systematic variance in reading comprehension should be explained by listening comprehension and word recognition. This was found in Study 1. In Study 2, however, syntactic skills continued to predict variance in the reading comprehension residuals, suggesting that the true/false test of reading comprehension makes demands on comprehenders' syntactic skills over and above those made by the test of listening comprehension.

It has previously been argued that the use of prosodic features in spoken language, such as stress and intonation, greatly facilitates the detection of the syntactic structure of the discourse (Rubin, 1980). The comprehension of a written text, it is argued, requires a much more sophisticated level of syntactic skill than does comprehension of the same text presented aurally (Adams, 1980). The results presented here provide empirical support to these arguments.

The second set of regression analyses also showed that, when using true/false tests, listening comprehension was predicted by non-verbal ability only, after controlling for reading comprehension. In Study 1, when standardised comprehension measures were used, listening comprehension had continued to be predicted by inferencing ability and vocabulary knowledge. Although the results differ in the two studies, what is interesting is that in both studies listening comprehension appears to make demands on comprehenders which are beyond those shared with reading comprehension. In

neither study, therefore, is there evidence that listening comprehension involves exactly the same processes as reading comprehension.

3. How might reliabilities of true/false tests affect findings?

Having discussed the results of the regression analyses, it is important to point out that the split-half reliabilities of the d-prime scores obtained on the true/false tests of comprehension were not particularly high. This suggests that the tasks lacked sensitivity. This is probably due to the fact that there were an insufficient number of items from which meaningful data could be obtained. Royer (2001) has pointed out that, when using the yes/no format of the Sentence Verification Technique, reliability increases with number of test items. He has found that, when children are asked to respond to 48 test sentences, reliability is in the range .5 to .6, as in the current study. However, when a 96-item test is administered, reliability is much improved at .8 to .9.

The poor reliability of the true/false tests used in this study means that results obtained from their use need to be interpreted with care.

5.6. Summary of Studies 1 and 2

Findings from Study 1 suggested that Listening Comprehension Test Series listening comprehension made demands on comprehenders over and above those shared with NARA-II reading comprehension. Of particular interest was the finding that listening comprehension appeared to make extra demands on children's inferencing ability. It was speculated that this finding might simply reflect the demands made by both the listening comprehension test and the inferencing test on children's memory for text. This explanation was tested in Study 2 when true/false tests of reading comprehension and listening comprehension were used that made similar demands on children's memory. Using these tests, it was found that comprehension in neither modality was predicted by inferencing ability. This suggested that the importance of inferencing in the standardised test of listening comprehension used in Study 1 was not due solely to the shared memory demands of the two tasks. Nevertheless, results from Study 2 have to be interpreted with care. The true/false tests of comprehension used have not been standardised and do not have good reliability, suggesting that they lack sensitivity.

Study 3 was conducted, therefore, to examine directly the relative contributions of memory for explicit information and inferencing ability in performance on the Listening Comprehension Test Series.

Chapter 6.

Study 3: Memory for explicit information and inference generation in listening comprehension

6.1. Introduction

Findings from Study 1 suggested that performance on the Listening Comprehension Test Series was strongly related to children's inferencing ability, having controlled for a variety of cognitive-linguistic factors. However, it is possible that children performing poorly on both the test of listening comprehension and the test of inference generation do so because they have a poor memory for text that they have heard. In Study 1 this possibility was not controlled for.

As discussed in Section 3.4, investigations of the role of memory for text in reading comprehension in children have proved inconclusive. However, there is no support for the idea that poor reading comprehenders fail to make inferences solely because they have poor memory for the text. Some studies have found that good and poor reading comprehenders differ on inferencing skill despite not differing on memory for literal information contained in a text (Cain & Oakhill, 1999; Oakhill, 1982; Oakhill, Cain, & Bryant, 2003). Other studies have found that good reading comprehenders perform better than poor reading comprehenders on questions requiring literal memory for the text as well as on questions requiring inference generation (Cain, Oakhill, Barnes, & Bryant, 2001; Oakhill, 1984). Cain et al. (2001) found, however, that differences between good and poor reading comprehenders in literal memory for the text did not fully account for their differences in inferencing ability. Similarly, Oakhill (1984) found that when the text was made available to poor comprehenders they were able to answer questions requiring them to extract literal information from the text. However, their problems with questions requiring inference generation remained. Thus, she concluded that poor comprehenders' problems with inferencing could not be due simply to poor memory for the literal content of the text.

It could be argued that it is not surprising that studies investigating literal memory for text and reading comprehension have not found conclusive evidence of a relationship

between the two. In all but two of the studies mentioned above (Oakhill, 1982; Oakhill, 1984) the groups of good and poor comprehenders were selected on the basis of their performance on a version of the NARA (the NARA-II) which allows children to look back at the text to answer questions. The demands on children's ability to remember explicit information from the text are therefore low and children's performance on this test will not necessarily depend on this skill. It seems likely that performance on the Listening Comprehension Test Series makes many more demands on children's ability to remember text than the NARA-II. Similarly, because the test of inferencing was administered aurally, it is likely that this too makes demands on children's ability to remember text. Barnes, Dennis, and Haefele-Kalvaitis (1996) found that when 6- to 15-year-olds heard stories, their performance on questions requiring them to generate knowledge-based inferences necessary for text coherence was predicted by their ability to answer questions requiring them to remember literal information in the text after controlling for age. It is therefore quite possible that children who perform well on both the Listening Comprehension Test Series and the test of inferencing used in Study 1 do so simply because they have a good memory for text that they have heard.

Whilst it is important to acknowledge the fact that memory for literal information potentially plays a role in performance on tests of both listening comprehension and inferencing ability, it should not be assumed that the relationship between memory for explicit information and inferencing is straightforward. Whilst it has been argued that good memory for literal information may be necessary for inferencing to occur (Barnes et al., 1996), it has also been argued that inferencing which occurs during encoding may itself potentially reinforce the memory representation of the literal propositions within a text (Perfetti, Landi, & Oakhill, 2005). In other words, children who generate inferences to form an elaborate and coherent representation of the text may, as a result, have a better memory for its explicit content.

A considerable amount of research evidence supports this argument. In their seminal work in the area, Bransford and Johnson (1972) suggested that, in a free recall task, comprehenders remembered few idea units from an apparently meaningless text that they had heard. Recall improved when participants were first given a title (e.g. "Washing Clothes") or a picture which provided a context for the text. When

participants understood the context of the material, they were able to make inferences which related the material they were hearing to their general knowledge and this improved their comprehension of the material and their memory for it. Other early research conducted by Paris and Upton (1976) suggested that inferencing during encoding improved memory for text in children. They found that 20 minutes after children had heard a short story, the best predictor of the number of idea units recalled was the number of inferencing questions that had been answered correctly when the story had first been heard. Similarly, Oakhill (1984) concluded that good reading comprehenders engaged in more inferential processing than poor reading comprehenders and that this improved their memory for literal information contained in the text. This conclusion was based on Oakhill's finding that, when the text could not be referred to, poor reading comprehenders were worse than good reading comprehenders at answering both questions tapping literal information and those requiring inference generation but, when the text could be referred to, poor comprehenders were worse than good comprehenders at answering questions requiring inferencing skills only.

There is also evidence that encouraging inferencing during encoding improves memory for the literal content of a text. Using a think-aloud procedure, which required participants to say what they were thinking about at the end of every sentence, van den Broek, Lorch, Linderholm, and Gustafson (2001) found that readers who were told that they would be examined on a text produced, as they were reading, more inferences to improve text coherence than participants who were told to read the text for enjoyment only. Participants were asked to complete a free recall task and duplicate the text as closely as possible. Those in the exam condition remembered more than those in the enjoyment condition. The authors argued that those in the exam condition had adopted higher "standards of coherence" when reading the text which had improved their understanding and memory for it. Similarly, Laing and Kamhi (2002) found that when children were asked to listen to stories using a think-aloud procedure, their performance in a recall task was higher than when they listened to the stories straight through. The authors argued that this was because the think-aloud procedure encouraged the generation of inferences and the creation of more coherent text representations.

Long and colleagues (Long & Prat, 2002; Long, Wilson, Hurley, & Prat, 2006) have provided further evidence that different comprehenders tend to construct qualitatively different types of representation of a text which lead to different memory experiences of it. In a series of experiments, memory for literal information presented in a text was assessed using recognition measures. Readers who had high levels of knowledge about the topic of a text were more likely to report consciously recollecting a sentence from the text than readers with little relevant background knowledge. Long, Johns, and Jonathan (submitted) argue that this pattern of performance reflects the fact that high knowledge readers are able to use background knowledge to engage in associative processes when reading. They build “networked” representations in which ideas are linked across the text and with existing knowledge. Low knowledge readers are unable to make connections between the text and their previous knowledge and their representation of the material is therefore “list-like” and unable to support the conscious recollection of test items.

The research presented here suggests that an individual’s tendency to generate inferences itself affects their ability to remember material that is explicitly stated in a text. However, not all studies have found this. As mentioned previously, a number of studies have found that good and poor reading comprehenders do not differ on memory for literal information contained in a text but do differ on inferencing skill (Cain & Oakhill, 1999; Oakhill, 1982; Oakhill, Cain, & Bryant, 2003). This finding is hard to explain if inferencing during reading leads to the development of stronger memory representations of literal information. Furthermore, Omanson, Warren, and Trabasso (1978) found that manipulations which enhanced the number of inferences made by children did not affect their ability to recall explicit information. In their study, children aged 5 and 8 listened to one of three versions of a story. In one, no mention was made of the protagonist’s motives whilst, in the others, the protagonist had motives that were either socially desirable or socially undesirable. After hearing the story, children freely recalled what they had heard before they were asked questions requiring the generation of inferences. It was found that when children were provided with information about the protagonist’s motives their ability to answer the inferencing questions improved. However, their ability to recall material that was semantically equivalent to the propositions that they had heard was unaffected. This

led Omanson et al. to conclude that memory for explicit text is not enhanced through inferencing.

6.2. Summary of aims

The aim of Study 3 was to clarify the relationship between memory for explicitly stated information, inferencing ability and performance on the Listening Comprehension Test Series.

In Study 1, performance on the Listening Comprehension Test Series was found to be predicted by vocabulary, non-verbal intelligence and inferencing skill. The role of memory for explicit information, however, was not explored. Thus, the first question addressed in Study 3 was whether or not memory for explicit information explained variance in listening comprehension after controlling for vocabulary and non-verbal intelligence. Although findings from the research literature are mixed concerning the role of memory for explicit information in reading comprehension, it was predicted that memory for literal material would predict listening comprehension.

Although Study 1 found inferencing ability to be a predictor of performance on the Listening Comprehension Test Series, the possibility that children performing poorly on both the test of listening comprehension and the test of inference generation did so because they had a poor memory for text had not been controlled for. Thus, the second question addressed in Study 3 was whether or not the ability to answer questions requiring inference generation continued to explain unique and significant variance in listening comprehension after controlling for memory for literal material. It was predicted that if the relationship between listening comprehension and inferencing was not due solely to the shared memory demands of the two tasks, inferencing would continue to explain variance in listening comprehension having controlled for memory for explicit information.

The third question addressed was whether or not memory for literal material in a text explained variance in listening comprehension after controlling for inferencing ability. If children with better inferencing skill construct more elaborate representations of text resulting in better access to explicit material, it would be expected that memory

for explicit information would not explain additional variance in listening comprehension having controlled for inferencing skill.

6.3. Method

Design

Study 1 suggested that vocabulary and non-verbal IQ were unique predictors of performance in the Listening Comprehension Test Series. These variables were, therefore, included as control variables, as was age. Two hierarchical regression analyses were conducted in which performance on the Listening Comprehension Test Series was the criterion variable. In the first analysis, control variables were entered on the first step, memory for literal information was entered on the second step, and the ability to answer questions requiring inference generation was entered on the third step. In the second analysis, steps two and three were reversed.

Participants

A new sample of children from two Calderdale primary schools took part in this study. According to the most recently published OFSTED reports, both schools were attended predominantly by White British children with a very small number of children from other ethnic backgrounds. Both schools had a lower than average number of children on free school meals. One of the schools had a higher than average number of children with a Statement of Special Educational Needs. Both schools followed the National Curriculum.

As in the previous studies, this study aimed to investigate the predictors of listening comprehension across a range of abilities, typical of those that might be found in a primary school classroom. Thus, the sample was again relatively unselected: whole year groups were assessed. The only children who were not included were those with a statement of Special Educational Needs due to learning difficulties. Children who needed to wear glasses to correct their eyesight were asked to do so whilst being assessed on any measure using visual stimuli.

As in the previous studies, children were assessed either in the last term of Year 3 or in the first term of Year 4. The final sample included 79 children, 48 of whom were in Year 3 at the time of testing, whilst 31 were in Year 4. The age-range of the children was 7 years, 10 months to 9 years, 0 months.

Study variables and measures

Each child was administered a battery of tests. This included the test of listening comprehension used in Study 1 as well as tests of the variables identified as its predictors, vocabulary, non-verbal IQ and inference generation ability. A test of memory for explicitly stated material was also administered. A list of the tests used is given below, as well as a description of the test of inferencing and memory for literal information.

Listening comprehension

As in Study 1, this was assessed using the Listening Comprehension Test Series (Level C) (LCTS: Hagues, Siddiqui, & Merwood, 1999).

The test was administered to the whole class according to the guidelines given in the test manual.

Receptive vocabulary

As in Studies 1 and 2, this was assessed using the second edition of the standardised British Picture Vocabulary Scale (BPVS-II: Dunn, Dunn, Whetton, & Burley, 1997).

Children were assessed individually according to the instructions given in the test manual.

Non-verbal IQ

As in Studies 1 and 2, this was assessed using Raven's Coloured Progressive Matrices (CPM: Raven, Raven, & Court, 1998).

Children were assessed individually according to the instructions given in the test manual.

Off-line inference generation and memory for explicitly stated information

These variables were assessed using the task devised by Oakhill (1984). In Study 1, Oakhill's task was adapted so that children were asked only those questions requiring them to make knowledge-based, gap-filling inferences. However, in the original version of Oakhill's test, children also had to answer questions which required them to remember information stated explicitly in the text. In Study 3, these questions were re-introduced into the test.

Children were assessed individually. They listened to the same four short passages which were used in the test of inference generation in Study 1. Each passage was followed by eight questions. The answers to four of these questions were stated explicitly in the text whilst the remaining four questions required children to generate knowledge-based inferences. All children heard the questions in the same order. As in previous research (Cain, Oakhill, Barnes, & Bryant, 2001), the order of the questions was chosen to reflect the order in which information was presented in the text. Furthermore, for each story, questions alternated between those requiring inference generation and those requiring memory for explicit information, as it has been argued that primacy or recency advantages for one type of question should be avoided (Paris & Upton, 1976). An example of a story from the test is shown in Table 20 and a copy of the full test is given in Appendix 6.

Table 20: Example story from the test of inferencing ability and memory for explicit information

Tim and the Biscuit Tin

Tim waited until he was alone in the house. The only sound he could hear was his father's axe on the logs in the shed. Tim looked in all the rooms again, to make sure his mother was not there. Then he pushed a chair over to the sink which was full of dishes. By climbing onto the edge of the sink, he could just reach the biscuit tin. The tin was behind the sugar. Tim stretched until his fingers could lift up the lid. Just as he reached inside, the door swung open and there stood his little sister.

1. Why did Tim want to be alone in the house? (Inferencing) *So he could steal/get/eat the biscuits. So he wouldn't be caught/told off*
2. Where was Tim's father? (Memory) *In the shed*
3. What was Tim's father doing? (Inferencing) *Chopping/cutting logs/wood*
4. What was in the sink? (Memory) *Dishes/bowls/washing-up*
5. How did Tim climb onto the sink? (Inferencing) *By pushing a chair over to the sink and standing on it. By using/standing on a chair*
6. Where was the biscuit tin? (Memory) *Behind the sugar*
7. What room was Tim in? (Inferencing) *Kitchen*
8. What happened as Tim reached inside the tin? (Memory) *His little sister came in/door swung open*

Children's responses to the questions were noted for later scoring. As in Study 1, the inferencing questions were only scored as correct if children had demonstrated that they had spontaneously gone beyond the information that they had been given to draw an appropriate inference. Memory questions were scored as correct if children gave the exact wording of the text or used close synonyms which retained the meaning of the original material (e.g. "washing-up" was accepted for "dishes" in example above). This approach has been used in previous research (Paris & Lindauer, 1976). A list of responses accepted as correct is given in Appendix 6. A second rater scored 40% of the sample using this rubric and inter-rater agreement was found to be very high on both inferencing questions (agreement on 97.7% of items) and memory questions (99.4%).

Split-half reliability for the inferencing test (Spearman-Brown correction) was calculated as .61, whilst for the memory test it was calculated as .84.

Any concerns that performance on the test is constrained by a child's general knowledge should be ameliorated by the fact that vocabulary, a close correlate of general knowledge (Cunningham & Stanovich, 1991; Ransby & Swanson, 2003), was also assessed in this study. Therefore if a relationship between knowledge-based inferencing and comprehension is identified which is independent of the relationship between vocabulary and comprehension, it can be assumed that the ability to use knowledge to generate inferences is itself involved in comprehension.

Procedure

The tests of vocabulary, non-verbal intelligence and inferencing and memory for explicit information were carried out with children individually in a quiet area of the school. They were all administered in the same session which lasted approximately 30/40 minutes.

The Listening Comprehension Test Series (Level C) was administered to the whole class over two half-hour sessions, normally separated by about a week. The first whole class session was held after each child had completed their individual session, so that all the children felt at ease with the researcher.

6.4. Results

Sample size

An *a priori* power analysis was conducted to determine the sample size required to test the multiple correlation. A medium-sized relationship between the predictors and the criterion variable was assumed ($f^2 = .15$). Results suggested that a sample size of 92 was required. Similarly, a minimum sample size of 90 is suggested by Tabachnick and Fidell (2007) for testing a multiple correlation given five predictor variables ($N > 50+40$). The actual sample size obtained ($N = 79$), therefore, fell a little short of the value suggested for testing the multiple correlation.

There was a very small amount of missing data (0.2%) which was dealt with by excluding cases listwise from the analysis.

Data preparation

The data was subjected to rigorous data screening as hierarchical regression techniques were to be used in the analysis.

Investigation of univariate distributions

Initial analyses were conducted to identify univariate outliers and atypical distributions. No cases were identified as outliers, i.e. no case had a score on any variable that was greater than 3.29 standard deviations above or below the mean score for that variable.

Visual inspection of histograms and expected normal probability plots for each variable suggested little deviation from normality. Values for skew and kurtosis were obtained and were found to be between +1 and -1 for each variable except for age, which had a value for kurtosis which was < -1 . Following procedures outlined by Tabachnick and Fidell (2007) and adopting the recommended conservative alpha level (.01), this kurtosis was not found to be significant ($z = 1.942, p > .01$).

Investigation of multivariate distributions and residuals

To screen the data for multivariate outliers and to test assumptions of normality, linearity and homoscedasticity of residuals, an initial standard multiple regression analysis was run with performance on the Listening Comprehension Test Series as criterion variable and age, vocabulary, non-verbal intelligence, memory for explicit information and inferencing ability as predictor variables.

The Mahalanobis distance for each case was obtained and evaluated using the chi-square distribution. With the use of a $p < .001$ criterion, none of the cases had a value of critical chi square in excess of 20.515 (5 IV's), i.e. no multivariate outliers were detected. Cook's distance was also requested for each case. No cases were found to possess a Cook's distance greater than 1, i.e. no case appeared to exert undue influence on the results.

Outliers in the solution, cases for which scores on the DV were poorly fit by the regression equation, were requested. No cases were found to have a standardised residual which was greater than +3 or lower than -3, i.e. no outliers in the solution were identified.

Normal probability plots and scatterplots of predicted values of the DV against standardised residuals were obtained to examine whether or not there were major deviations from normality, linearity and homoscedasticity. No deviations were identified.

Descriptive statistics

Because ages of participants were noted, age could be independently controlled for in the hierarchical regression analyses that follow. Thus, raw scores were used for all measures. Table 21 shows the descriptive statistics obtained for each measure using these raw scores.

Table 21: Means and standard deviations for all Study 3 measures (raw scores)

Measure	N	Mean	SD	Skew	Kurtosis
LCTS comprehension	78	24.51	5.21	-.567	.217
Age	79	100.86	3.99	-.081	-1.070
BPVS-II	79	88.37	10.844	.117	.569
Raven's CPM	79	27.54	4.063	-.284	-.349
Inferencing	79	8.94	2.705	-.183	-.358
Explicit memory	79	9.78	3.608	-.323	-.582

Key: LCTS comprehension – Listening comprehension from the Listening Comprehension Test Series; Age – chronological age in months; BPVS-II – British Picture Vocabulary Scales (Second Edition); Raven's CPM – Raven's Coloured Progressive Matrices (1998 Edition); Inferencing – Correct inference responses (Max = 16); Explicit memory – Correct memory responses (Max = 16).

Hierarchical regression analyses

Two hierarchical regression analyses were conducted. The purpose of the first was to ascertain whether inferencing ability continued to explain variance in listening

comprehension having controlled for age, vocabulary, non-verbal intelligence and memory for literal information. The purpose of the second was to find out whether memory for literal information itself made a unique contribution to listening comprehension having controlled for age, vocabulary, non-verbal intelligence and inferencing skill.

Bivariate correlations (two-tailed) between the variables were firstly obtained to check for evidence of multicollinearity between the predictor variables. Results are shown in Table 22. Because of the number of correlations, a Bonferroni correction was applied, giving a criterion for significance of $p < .0033$ ($.05/15$).

Table 22: Correlations between Study 3 measures

Variable	1	2	3	4	5	6
1. LCTS comprehension	-	.155	.630**	.410**	.588**	.524**
2. Age		-	.267	.221	.008	.089
3. BPVS-II			-	.512**	.493**	.434**
4. Raven's CPM				-	.117	.174
5. Inferencing ability					-	.634**
6. Explicit memory						-

** $p < .001$

* $p < .003$

Investigation of the correlation matrix shown in Table 22 suggested no evidence of multicollinearity between the predictor variables. None of the correlation coefficients for relationships between predictor variables exceeded the value of .80.

Results of the hierarchical regression analyses are shown in Table 23. Listening comprehension, as measured by performance on the Listening Comprehension Test Series, was the criterion variable. In both analyses, age, vocabulary (BPVS-II) and non-verbal intelligence (Raven's CPM) were entered on the first step. In the first analysis, memory for explicit information was entered on the second step and

inferencing ability was entered on the third step. In the second analysis this order was reversed.

Table 23: Hierarchical regression analyses with LCTS listening comprehension as the criterion variable

Variable	B	SE B	Beta	t	Sig t	ΔR^2	Sig ΔR^2
Step 1							
Age	-.034	.122	-.026	-.275	.784		
BPVS-II	.275	.051	.574	5.420	<.001		
Raven's CPM	.155	.133	.122	1.165	.248	.408	<.001
Step 2							
Age	-.024	.114	-.018	-.207	.836		
BPVS-II	.203	.052	.425	3.916	<.001		
Raven's CPM	.180	.125	.142	1.442	.153		
Explicit memory	.454	.134	.316	3.395	.001	.081	.001
Step 3							
Age	.020	.110	.015	.180	.857		
BPVS-II	.147	.053	.307	2.752	.007		
Raven's CPM	.236	.121	.185	1.948	.055		
Explicit memory	.226	.151	.157	1.490	.141		
Inferencing	.604	.214	.315	2.819	.006	.051	.006
Step 2							
Age	.029	.111	.022	.259	.796		
BPVS-II	.157	.054	.327	2.925	.005		
Raven's CPM	.243	.122	.190	1.990	.050		
Inferencing	.775	.182	.404	4.249	<.001	.117	<.001
Step 3							
Age	.020	.110	.015	.180	.857		
BPVS-II	.147	.053	.307	2.752	.007		
Raven's CPM	.236	.121	.185	1.948	.055		
Inferencing	.604	.214	.315	2.819	.006		
Explicit memory	.226	.151	.157	1.490	.141	.014	.141

When age, vocabulary (BPVS-II) and non-verbal intelligence (Raven's CPM) were entered on the first step, they accounted for 40.8% of the variance in performance on

the Listening Comprehension Test Series, $F(3, 74) = 16.968, p < .001$. Only the regression coefficient for vocabulary differed significantly from zero at this step.

In the first analysis, memory for explicit information was entered into the regression model at Step 2 which resulted in the explanation of an additional 8.1% of the variance in listening comprehension, $F \text{ change}(1, 73) = 11.526, p = .001$. The total variance explained by the model as a whole at the end of Step 2 was 48.8%, $F(4, 73) = 17.418, p < .001$. Two regression coefficients differed significantly from zero at this step, vocabulary and memory for explicit information. At Step 3, inferencing ability was entered into the regression equation and an additional 5.1% of the variance was explained, $F \text{ change}(1, 72) = 7.949, p = .006$.

In the second analysis, inferencing ability was entered into the regression model at Step 2. This resulted in the explanation of an additional 11.7% of the variance in listening comprehension, $F \text{ change}(1, 73) = 18.052, p < .001$. The total variance explained by the model as a whole when inferencing had been entered at Step 2 was 52.5%, $F(4, 73) = 20.171, p < .001$. Two regression coefficients differed significantly from zero at this step, vocabulary and inferencing ability, and the unique contribution of non-verbal intelligence approached significance. When memory for explicit information was added to the regression model at Step 3, there was no significant increase in the variance in listening comprehension explained, $F \text{ change}(1, 72) = 2.219, p = .141$.

Obviously, for both analyses the final model was the same. The total variance explained was 53.9% (adjusted $R^2 = 50.7$), $F(5, 72) = 16.850, p < .001$. Only two variables were unique, significant predictors in the final model, vocabulary and inferencing ability, although the contribution of non-verbal intelligence approached significance.

These results show that the ability to answer questions requiring inference generation accounted for significant variance in listening comprehension even after controlling for memory for explicitly stated information. Recall of explicitly stated information, however, did not explain further variance in listening comprehension after the contribution made by inferencing ability had been taken into account.

6.5. Discussion

This study aimed to explore the relationship between inferencing skill, memory for explicit information and listening comprehension by addressing three questions. These were explored by conducting two hierarchical regression analyses. In both of these analyses the control variables of age, vocabulary and non-verbal intelligence were entered on the first step. In the first analysis, memory for literal information was entered on the second step and inferencing ability was entered on the third step. In the second analysis these last two steps were reversed.

The first question was whether or not memory for the explicit content of a text explains variance in listening comprehension having controlled for vocabulary and non-verbal intelligence. Results presented above show that, as predicted, when entered after the control variables, memory for literal information does predict unique variance in listening comprehension ability. Comparable results have been found in some studies of reading comprehension which find that children who are poor reading comprehenders have a poorer literal memory for a text than children who are good reading comprehenders (Cain, Oakhill, Barnes, & Bryant, 2001; Oakhill, 1984).

The second question was whether or not the relationship identified in Study 1 between performance on the Listening Comprehension Test Series and performance on the test of inferencing simply reflects the shared memory demands of the two tasks. This was addressed by investigating whether the ability to answer questions requiring inference generation continues to explain unique and significant variance in listening comprehension after controlling for memory for literal material. Results presented above show that inferencing ability does continue to explain variance in listening comprehension having controlled for memory for explicit information. In other words, the relationship between listening comprehension and inferencing ability is not wholly explained by variance that inferencing skill shares with the ability to remember literal information from a text. This is consistent with previous findings from the reading comprehension literature which have found that differences between the inferencing skills of good and poor comprehenders cannot be fully explained by differences in

memory for explicit information (Cain, Oakhill, Barnes, & Bryant, 2001; Oakhill, 1984).

The third question was whether or not memory for literal material in a text explained variance in listening comprehension after controlling for inferencing ability. Results presented above show that memory for explicit information does not explain additional variance in listening comprehension having controlled for inferencing skill. The relationship of listening comprehension with the ability to recall explicitly stated information is mediated entirely by its relationship with the ability to answer questions requiring the generation of inferences.

The results presented here suggest that individual differences in memory for explicit information cannot explain individual differences in inferencing skill. Children who demonstrate similar ability to recall explicitly stated information may differ in their ability to answer questions requiring inferencing. However, individual differences in inferencing ability appear to explain individual differences in the ability to remember explicit information. A child with good inferencing skills will necessarily have good memory for explicit text information.

These findings are consistent with the view, described in the introduction to this chapter, that children who engage in inferential processing as they listen to a text create a semantically rich, coherent representation of the material which enables them to access both explicit and inferred information at test (Oakhill, 1984; Paris & Upton, 1976; Perfetti, Landi, & Oakhill, 2005). Children who do not engage in such inferential processing at encoding may still store propositions from the text in memory, but their representations will be “list-like” and, although they may be able to recall explicit information, inferred information will be unavailable to them. These children may be able to recall the same amount of literal information as children with more “networked” representations but they will not be able to answer the same number of inferencing questions.

Whilst the findings of Study 3 are consistent with the view that inferencing during encoding aids performance on questions requiring both inference generation and literal memory for text, it is important to acknowledge that an alternative explanation

is also possible. There is no concrete evidence that the children in this study make inferences as they listen to the text. Some authors have argued that very few inferences are made at encoding and that most are made strategically at test (McKoon & Ratcliff, 1992). According to this view, having access to a memory representation of the explicit information at test is a necessary condition for inferencing to occur but not a sufficient one. The best comprehenders are those children who can remember explicitly stated material and generate elaborative inferences at test. Children who recall the same amount of literal information will not necessarily possess the same inferencing skill. Therefore, individual differences in memory for explicit information cannot explain individual differences in the ability to answer inferencing questions. However, because memory for explicit information is a prerequisite for inferencing to occur, the test of inferencing itself assesses children's memory skills. Hence, individual differences in performance on the inferencing test would explain individual differences in the ability to remember explicit information. As can be seen, therefore, this alternative explanation fully accounts for the results found.

6.6. Summary of Study 3

The results of this study show that the importance of inferencing skill in performance on the Listening Comprehension Test Series is not due entirely to the shared memory demands of the two tasks. Whilst this is an important finding, it is not clear whether inferencing generally occurs at encoding or at test, and the exact nature of the relationship between inferencing and memory for explicit information is still unclear. The reason why it is impossible to distinguish between the two possible explanations of the findings given above is that the test of inferencing used was an off-line test, i.e. it was administered after rather than during encoding. Off-line tasks cannot discriminate between inferences drawn during and after presentation of material. The aim of Study 4, therefore, was to explore the relationship of listening comprehension with the on-line generation of inferences.

Chapter 7.

Study 4: On-line inferencing and listening comprehension

7.1. Introduction

Findings from Study 3 suggested that the relationship between performance on the Listening Comprehension Test Series and performance on an off-line test of inferencing ability did not arise solely because of the shared memory demands of the two tasks. Having a memory for explicitly stated propositions in a text does not appear to guarantee inferencing success. It was also found that children who can inference successfully necessarily have good literal memory for the text. Two potential explanations for these findings were proposed. The first explanation was that children who were good comprehenders engaged in inferencing whilst listening to the text to form a rich, elaborated, “networked” representation of the material in the text. Children who were poor comprehenders did not engage so extensively in inferential processing and their representations were more “list-like”. The children with more “networked” representations were more able than those with “list-like” representations to answer questions requiring inference generation and questions requiring memory for explicitly stated propositions. The second explanation, equally compatible with the data obtained, was that children did not generate the inferences until they were asked the questions. A good representation of the literal propositions of the text was, according to this explanation, a necessary but insufficient condition for inferencing to occur. Better comprehenders were simply better at inferencing at test. For example, in one of the stories used in the inferencing test, children heard the sentence “It was the flapping of wings” and were asked “What creature was making the noise?” It was not clear whether children giving the correct answer had actually inferred the creature was a bird as soon as they had heard the relevant sentence in the story (as suggested by the first explanation given above) or whether they had inferred it only when they were asked the question (as suggested by the second explanation). Singer (1976) referred to these possibilities as “Inference-on-Input” and “Inference-Later” respectively. The fact that the test of inferencing used was an off-line test meant that it gave no information as to when the inferences were generated.

It was important to distinguish between these two explanations. Performance on the inferencing test was shown to be an important predictor of listening comprehension in Studies 1 and 3. To understand more fully the nature of the demands that listening comprehension makes on inferencing skill, Study 4 was conducted to investigate whether or not listening comprehension is associated with inferencing that occurs at encoding, as material is heard.

As mentioned in Chapter 3, there is a huge amount of debate within the reading comprehension literature regarding the amount of inferential processing that occurs “on-line” during encoding. Whilst it is generally accepted that on-line inferencing occurs in order to maintain text coherence, there is little consensus as to what this actually means. Proponents of the “minimalist” view” argue that coherence is only important at a local level and that minimal inferencing, such as pronoun resolution, occurs during encoding. According to this view, most inferencing occurs at test and is the result of retrieval operations which act on a minimalist representation of the text in memory (McKoon & Ratcliff, 1992). Other authors adopt a “constructionist” view and argue that global coherence is striven for as well as local coherence (Graesser, Singer, & Trabasso, 1994). Graesser et al. (1994) have identified 13 types of inference, six of which they believe to be generated routinely during encoding. Such inferences include, for example, the goals and motivations of characters in a story and the causal relationships between events. To illustrate the difference between the minimalist and constructionist positions, McKoon and Ratcliff (1992) refer to the seminal work of Bransford, Barclay, and Franks (1972). These early researchers, who adopted a strong constructionist approach, argued that when participants heard “Three turtles rested on a floating log, and a fish swam beneath them”, they automatically encoded the inference that the fish swam beneath the log as part of their situation model of the text. The minimalist position is that this inference would not be automatically encoded because it would not be necessary for local text coherence. Instead, participants would incorrectly “recognise” the foil “Three turtles rested on a floating log, and a fish swam beneath it” due to backward processes occurring on hearing the foil. Recently, attempts have been made to integrate the two positions. For example, Long and Lea (2005) have argued that the “search after meaning” requires both automatic, passive processes advocated by the minimalist approach and a strategic effortful evaluation of activated information and its integration into the

situation model. Van den Broek, Rapp, and Kendeou (2005) invoke the concept of “standards of coherence” which vary both between readers and within readers according to their reading goals. They argue that information activated automatically is evaluated with respect to the comprehenders’ standards of coherence. If activated information does not satisfy these standards, more effortful, strategic processes will be employed.

Several authors have used findings from studies involving off-line tests to argue that knowledge-based inferences are generated on-line. Paris and Lindauer (1976) compared the ability of a cue to facilitate the recall of a sentence when the cue was either an explicitly stated or implied instrument of the sentence. For example, they compared the ability of the word “broom” to cue recall of the sentences “Her friend swept the kitchen floor” and “Her friend swept the kitchen floor with a broom”. Paris and Lindauer found that for 11-12 year-olds, but not 6-7 year-olds, the cue was equally effective at prompting recall whether it had been stated directly or implied. Similarly, Paris, Lindauer, and Cox (1977) found that, for 11-12 year-olds, but not 6-7 year-olds, a cue which was a consequence of a sentence facilitated sentence recall whether the consequence had been explicitly stated in the sentence or implied. The authors concluded that the older children were spontaneously generating the inferred instruments and consequences at encoding. Other authors, however, have argued that this conclusion is incorrect. Corbett and Doshier (1978), for example, found that an instrument cued recall for a sentence even when it was unlikely that it had been inferred at encoding because another instrument had been mentioned explicitly. These authors argued that recall is facilitated by a backward association between the cue and the sentence and that inferences are made not on-line but at recall after hearing the cue.

The limitation with the use of off-line tests is that, although an attempt can be made to infer what is happening on-line from the results, this information is not provided directly. The direct investigation of the generation of inferences during encoding requires the use of on-line techniques which measure processing as it actually occurs. One such technique measures naming latencies to words. Using this technique, Potts, Keening, and Golding (1988) showed that words were primed by short passages which prompted them as inferred consequences. Furthermore, such priming only

occurred when the inferred consequences were necessary for coherence and provided an explanation for the content of the passages. Inferences such as these are referred to in the literature as “bridging” or “backwards” inferences. Potts et al. found that when the inferred consequences were not necessary to explain text events but were instead “forward” or “predictive” inferences they were not primed. For example, the word “broke” was primed by the passage “No longer able to control his anger, the husband threw the delicate porcelain vase against the wall. It cost him well over one hundred dollars to replace the vase.” In order for the two sentences to make sense, it was necessary for the comprehender to infer that the vase was broken. The word “broke” was not primed, however, following the passage “No longer able to control his anger, the husband threw the delicate porcelain vase against the wall. He had been feeling angry for weeks, but had refused to seek help”. In this case, whilst it is quite possible that the vase was broken, it is not necessary to infer this in order for the passage to make sense. This inference is more “optional”. Potts et al. concluded that inferences were generated on-line but only when they were necessary to make sense of preceding propositions.

A further on-line technique that has been used to investigate inferencing is the lexical decision task. In this task, participants have to decide whether a letter string is a real word or a nonsense word. Till, Mross, and Kintsch (1988) used this technique to investigate priming of words which were thematically-related to short passages. For example, participants read “The townspeople were amazed to find that all the buildings had collapsed except the mint. Obviously, it had been built to withstand natural disasters.” They then had to make a lexical decision to a target item which was either a thematically-appropriate inference, e.g. “earthquake”, or an inappropriate topic word, e.g. “breath”. By comparing decision latencies to words which were topical inferences and to words which were unrelated to the preceding propositions, Till et al. were able to demonstrate that priming to the inference words, relative to the inappropriate words, did occur, but only after a time of 1000msec had elapsed from the end of the passage.

The studies mentioned above have looked at the on-line generation of inferences in groups of mixed comprehension ability. Of particular interest to the current study, however, is research which has used on-line techniques to identify individual

differences in on-line inference generation. Long, Oppy, and Seely (1994) compared the performance of skilled and less skilled comprehenders on the task devised by Till et al.. They found that skilled comprehenders responded faster to appropriate than to inappropriate inference words when only 500ms had elapsed from the end of the passage. Less skilled comprehenders, however, did not respond faster to appropriate topic words even after 1000msec had elapsed. There was, however, evidence that the less skilled comprehenders did construct accurate sentence-level representations. There was also evidence that the less skilled comprehenders were able to make the correct knowledge-based inferences when asked explicitly to do so. The authors concluded that skilled comprehenders engaged in on-line knowledge-based inferential processing whereas less skilled readers did not. Long et al.'s findings were replicated by Hannon and Daneman (1998).

Very few studies have investigated any aspect of children's comprehension using on-line techniques. Nation and Snowling (1999) used an aurally administered lexical decision task to compare semantic priming in good and poor comprehenders. Similarly, they used a naming latency task to compare the reading of irregular words by good and poor comprehenders. Specifically, they investigated the extent to which irregular word reading was supported by semantics (Nation & Snowling, 1998b) and facilitated by sentence context (Nation & Snowling, 1998a). These studies suggest that on-line tests can be devised which are sensitive to individual differences in children's comprehension ability.

An investigation into children's on-line inference generation, using sentence reading time as a measure of on-line processing, has been conducted by Casteel (1993). He found that children spent longer reading a sentence when it was necessary to generate an inference to link the material in the sentence with previously presented text than when inference generation was unnecessary. For example, children read either "Amy and her friends had a slumber party. Amy's father told them a ghost story. The girls had never heard the story" or "Amy and her friends had a slumber party. Amy's father told them a ghost story. The story scared all of the girls". The time taken by children to read the sentence "The girls left the light on all night" was measured following each passage. Casteel found that children spent longer reading the test sentence when it followed the first passage than when it followed the second passage. He

hypothesised that this was because children reading the first passage generated the inference that the girls were scared when they read the test sentence in order to provide a causal link between material in the passage and the test sentence.

Evidence from Casteel's study suggests that children do generate knowledge-based inferences on-line whilst reading when such inferences are necessary to make causal links between text propositions. However, it does not reveal any information concerning individual differences in the tendency to generate inferences on-line. Furthermore, the study did not explore the relationship between on-line inferencing ability after taking into account other predictor variables known to be important in reading comprehension in children. To date, all studies of individual differences in children's comprehension which have considered inferencing ability alongside other predictor variables have used off-line measures of inferencing skill (Cain, Oakhill, & Bryant, 2004). Recently, it has been argued that studies of individual differences in comprehension in children should incorporate measures of the ability to generate inferences on-line (Kendeou, Bohn-Gettler, White, & van den Broek, 2008).

7.2. Summary of Aims

Findings from Studies 1 and 3 suggested that children's performance on an off-line test of inference generation was an important predictor of performance on the Listening Comprehension Test Series. The aim of Study 4 was to examine whether performance on the Listening Comprehension Test Series was related to the ability to generate on-line knowledge-based inferences. Based on findings from the reading comprehension literature, it was hypothesised that there would be a relationship between listening comprehension ability and the tendency to generate inferences whilst listening. Data was collected from the children taking part in Study 1 who had already been assessed on a large number of variables identified in the literature as being related to comprehension. This meant that, should a bivariate correlation between listening comprehension and on-line inferencing be identified, there would be the opportunity to investigate, through regression analysis, whether or not on-line inferencing made a unique contribution to listening comprehension having controlled for other cognitive-linguistic variables.

7.3. Method

Design

The study was correlational in design. A lexical decision task was used in which children heard sentences followed by a letter string and were asked to decide whether or not the letter string was a real word. For each child, an on-line inferencing score was calculated which reflected the difference in their response latencies to words which were thematically-related inferences of the preceding sentences and to words which were unrelated to the sentences. The inferencing score also took into account children's tendency to make semantic associations between target words and the words in the preceding sentences. The correlation between the on-line inferencing score and performance on the Listening Comprehension Test Series was obtained.

Participants

Data were collected from the same 126 children who took part in Studies 1 and 2.

Materials and Procedure

The lexical decision task was administered on the computer using E-prime software. Long et al.'s (1994) test has been modified by Spooner and Willis (in preparation) so that the materials are suitable for children and presentation of materials is auditory. The current study used materials which were adapted from those used by Spooner and Willis.

In total, children heard 32 sentences. Figure 4 illustrates the nature of the sentences, coherent and scrambled, and of the stimuli that followed them.

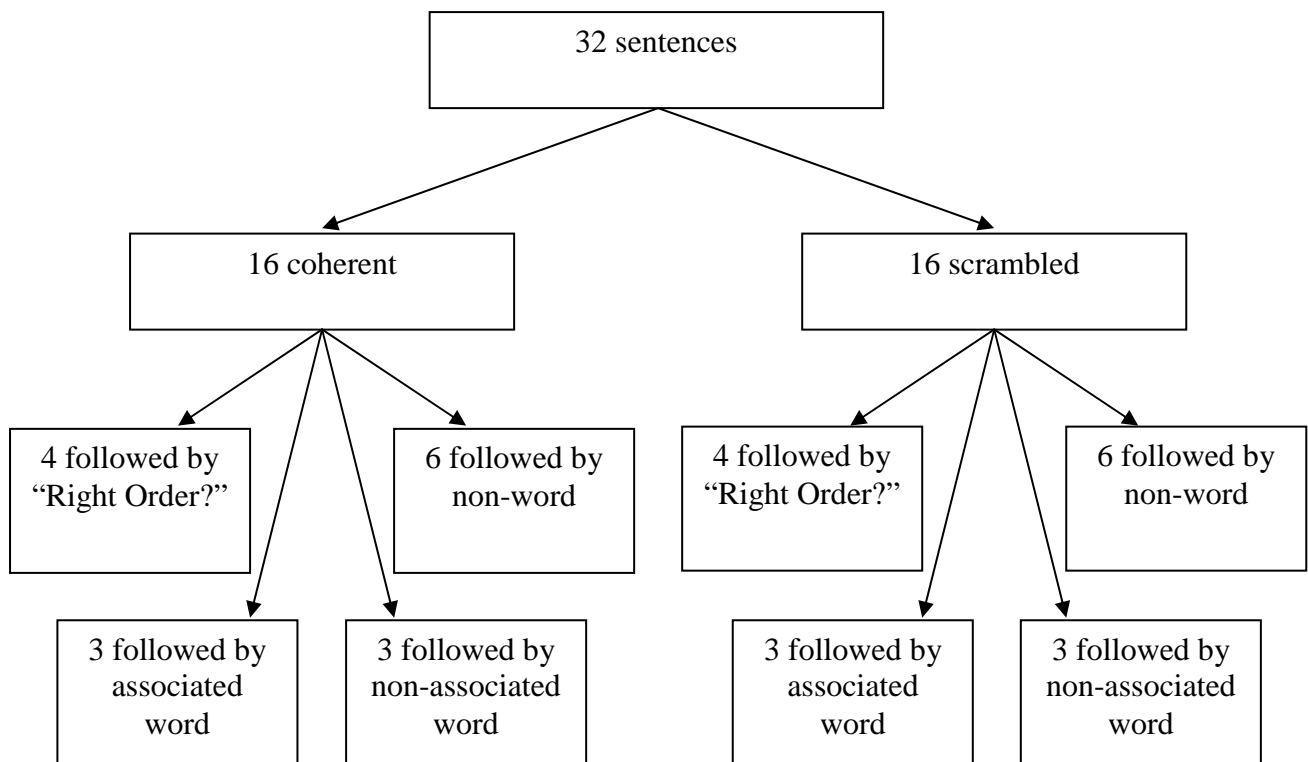


Figure 4: Nature of stimuli used in on-line inferencing test

Twelve of the coherent sentences and 12 of the scrambled sentences were followed by an aurally-presented letter-string. Each letter-string was accompanied by a visually-presented question mark which appeared on the screen. Children had to decide whether each letter string was a word (e.g. “dark”) or non-word (e.g. “lort”). To make their response, children used a computer mouse, held as a Sony Playstation controller. A tick had been attached to the left-hand mouse button and a cross had been attached to the right-hand mouse button. Children pressed the “tick”, if they thought the letter string was a word, or the “cross”, if they thought the letter-string was a non-word. The letter-strings following 6 of the coherent sentences and 6 of the scrambled sentences were words. A further 6 coherent sentences and 6 scrambled sentences were followed by non-words.

The 6 words following coherent sentences were presented in one of two conditions. Three words followed sentences which prompted the target words as knowledge-based, thematically-related inferences. For example, in one case, children heard “The

parents flew back and forth with worms for their babies” followed by the word “Birds”. A further three words followed sentences which had no association to the words at all. For example, children heard “She’d got the time wrong and no-one else was there yet” followed by the word “Milk”. If children did generate inferences on-line as they were listening, it would be expected that their response times to the words following sentences to which they were related would be faster than their response times to the words following non-associated sentences.

It has, however, been argued that, in order to demonstrate that inferencing is occurring, a lexical decision task needs to control for associative priming between the target word and the individual words to which it is related in the sentence (Potts et al., 1988). In other words, if response times to words following associated sentences are faster than response times to words following non-associated sentences, it is important to show that this is due to on-line processing involving knowledge of the situation described by the sentences and is not due simply to semantic relatedness of the target word and the sentence words.

This was the reason for the inclusion of the scrambled sentences. Three of the scrambled sentences were followed by words which provided a knowledge-based, thematic inference related to the situation which would have been described by the sentence had it been coherent. For example, the original sentence “He wriggled and wriggled but he just could not get out” was scrambled and presented to the children as “He wriggled and could but he just get not wriggled out”. This was followed by the target word “Stuck”. As the words in the sentence were scrambled, it was not possible for children to make an inference based on the situation described by the sentence. It was possible, however, that the word “stuck” could be primed by individual words within the scrambled sentence such as “wriggled”. In the second condition, three words followed scrambled sentences which had no association with the target words. For example, the original sentence “He drove too fast and I was trembling when he stopped” was scrambled and presented as “I stopped too fast and he was drove when he trembling”. This was followed by the target word “tree”. Any difference in response times to the words in these two conditions would have to be due to priming arising from semantic relatedness of the target and sentence words and could not be due to knowledge-based inferencing per se.

A summary of the conditions in which target words appeared is given in Table 24.

Table 24: Illustrative examples of the four conditions in which target words appeared

	<i>Target word associated with sentence</i>	<i>Target word not associated with sentence</i>
Coherent sentence	Sentence: The parents flew back and forth with worms for their babies. Target: Birds	Sentence: She'd got the time wrong and no-one else was there yet. Target: Milk
Scrambled sentence	Sentence: He wriggled and could but he just get not wriggled out. Target: Stuck	Sentence: I stopped too fast and he was drove when he trembling. Target: Tree

The presentation of words in the four conditions outlined above meant that it was possible to calculate, for each child, a score which reflected their tendency to generate inferences whilst taking into account their tendency to engage in associative priming. To obtain a score for the extent to which each child engaged in priming when associated words followed coherent sentences, the following calculation was made:

$$\text{Priming following coherent sentences} = \text{Mean RT (Non-associated/Coherent)} - \text{Mean RT (Associated/Coherent)}$$

This priming score was assumed to reflect priming resulting from the generation of inferences, related to the situation described by the sentence, and priming resulting from the semantic relatedness of words in the sentence and the target words. To obtain a score for the extent to which each child engaged in associative priming between sentence and target words, priming to associated words following scrambled sentences was calculated as follows:

$$\text{Priming following scrambled sentences} = \text{Mean RT (Non-associated/Scrambled)} - \text{Mean RT (Associated/Scrambled)}$$

The difference between these two scores, obtained for each child, gave an indication of their tendency to generate inferences after hearing a sentence whilst accounting for their tendency simply to activate words related to the sentence words but not, necessarily, to the whole meaning of the sentence. Therefore, the overall inferencing score for each child was calculated as follows:

$$\text{Overall inferencing score} = \text{Priming following coherent sentences} - \text{Priming following scrambled sentences}$$

As shown in Figure 4, six of the coherent sentences and six of the scrambled sentences were followed by non-words. For example, children heard the sentence “You can stroke the animals and walk in the fields” followed by the non-word “lote”. In addition, four of the coherent sentences and four of the scrambled sentences were verification sentences. These sentences were not followed by a letter string. Instead, they were followed by the question “Right order?”. The children had to indicate whether or not they thought the sentence was correctly ordered by pressing the tick or the cross on the mouse buttons. The inclusion of these sentences was designed to ensure that children attended to the sentences themselves and did not adopt a strategy of focussing exclusively on the letter strings.

At the beginning of the test, children were all given the same instructions. They were told that they were going to hear some sentences on the computer, some of which would be jumbled and some of which would be in the correct order. They were told that, after each sentence, one of two things would happen. They might hear the computer saying “Right Order?” If that happened they had to think about the sentence they had heard and decide whether or not it was in the right order. On the other hand, they might hear the computer saying something that sounded like a word which would either be a real word or a made-up word. Their job then was to decide whether or not the word was real. The children were told to try and respond as quickly as possible, but without making mistakes. The children had four practice trials before the test

started. If they still appeared to be struggling with the demands of the task, which was rarely the case, the four practice trials were repeated.

After the presentation of each sentence, there was an inter-stimulus interval of 600ms, before the presentation of the letter string or of the question “Right order?”. This interval was chosen as it has been found in studies of adult readers that the more skilled comprehenders make inferences at this ISI (Long et al., 1994). A timeline for the experiment is shown in Figure 5.

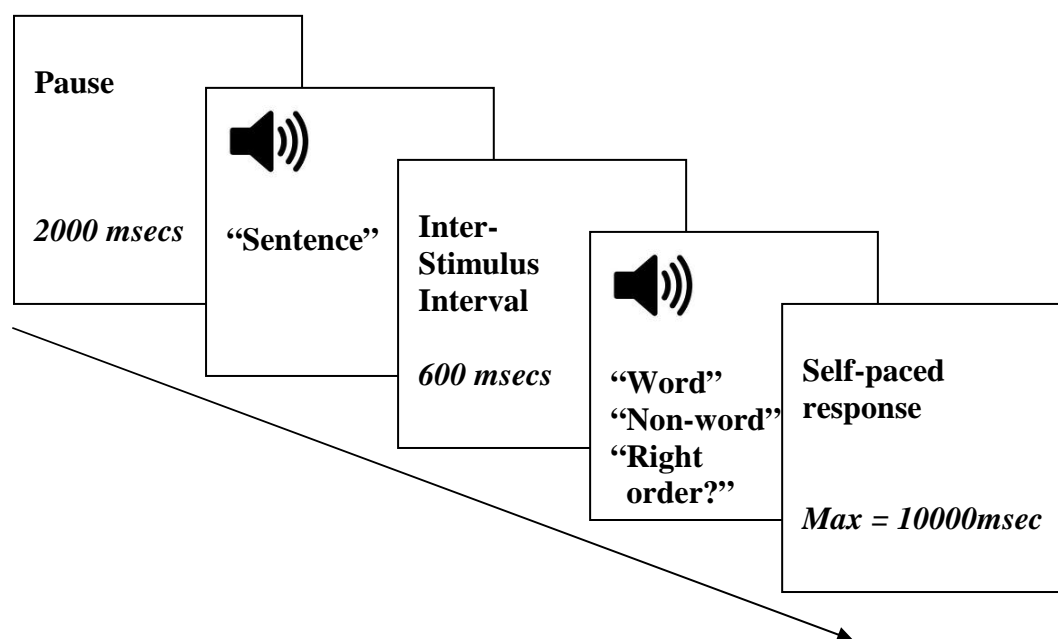


Figure 5: Experimental timeline

The order of presentation of the stimuli was randomised using the E-prime software. As the study utilised a correlational design, all children were presented with exactly the same stimuli in exactly the same random order. Appendix 7 shows the stimulus words and sentences heard by the children.

The materials were piloted on six Year 2 children. The purpose of the pilot was to ensure that the task was not too complex for young children to carry out. The children taking part in the pilot study demonstrated that they understood what they were

expected to do and that they were able to use the equipment appropriately to make their desired responses.

The accuracy and latency of all responses were recorded on the computer. Children were tested individually in a quiet area of the school. The test took approximately 10 minutes to administer.

Children taking part in this study had also been assessed on the Listening Comprehension Test Series (Hagues, Siddiqui, & Merwood, 1999).

7.4. Results

Data preparation

Lexical Decision Errors

Overall, there were few errors on the lexical decision task, most children performing at ceiling. Following McNamara and McDaniel (2004) who advocate an inclusion criterion for a participant of 70% accuracy, one child was removed from the analysis because their identification of words and non-words was incorrect on more than 33% of the trials. In line with previous studies, response latencies to words erroneously identified as non-words were not included in the analysis (Long, Oppy, & Seely, 1997; Long, Seely, & Oppy, 1999).

Outliers

Ratcliff (1993) describes reaction time outliers as reaction times that arise due to processes other than those of interest, for example inattention and daydreaming. Whilst it is obviously desirable to eliminate outliers from the data, their identification is not unambiguous. As Ratcliff points out, there is no clear distinction between the distribution of response times resulting from the processes of interest and the distribution of outlier response times. It is important, therefore, that the procedure used to eliminate outliers does not inadvertently result in the elimination of some of the data that is of value.

Outliers for each of the four conditions were identified separately. A very small number of exceptionally long reaction times ($> 7000\text{msec}$) were first eliminated. Examination of the distribution of reaction times to target items suggested that these reaction times did not seem to be connected to the rest of the distribution. For each condition, the mean response time to word items was then calculated and any value which was greater than three SD's above this was removed. Two percent of the data was removed by following this procedure.

Number of items per cell

In the current study, children responded to only three words in each of the four conditions. In situations where, due to the screening processes described above, any condition for any child now contained fewer than two items, that child was excluded from further analysis. Thus, the analysis reported here includes only children who had provided two or more accurate responses within the acceptable time-scale for each of the four study conditions. Overall, 5.5% of the cases were excluded. The number of children included in the analysis was 118.

Does the test detect priming due to inferencing?

Before investigating the relationship between the on-line inferencing score and performance on the Listening Comprehension Test Series, it was necessary to ascertain whether or not the test was detecting inference generation.

A 2-way repeated measures ANOVA was conducted with association (associated, non-associated) and sentence coherence (coherent, scrambled) as factors. It was expected that, if the test detected priming, reaction times to associated items would be faster than reaction times to non-associated items. Furthermore, if the priming detected by the test was due to children's generation of inferences rather than the formation of simple lexical associations between target items and words in preceding sentences, the difference between reaction times to associated and non-associated words would be greater in the coherent condition than in the scrambled condition. In other words, it was expected that, if the test was detecting priming due to inference generation, an interaction would be evident between association and coherence such that the effect of association on reaction time was greater in the coherent than in the scrambled condition.

Descriptive statistics were obtained for each of the four conditions and are shown in Table 25.

Table 25: Means and SD's for response latencies to words in each condition

<i>Condition</i>	<i>Mean RT</i>	<i>SD</i>
Associated/Coherent	944.14	347.52
Non-associated/Coherent	1194.76	420.79
Associated/Scrambled	1163.28	454.77
Non-associated/Scrambled	1243.83	474.90

As the table shows, in the coherent condition the mean response time to associated words was 250.62ms faster than the mean response time to non-associated words, whilst in the scrambled condition this difference was much less at 80.55ms. A two-way repeated-measures ANOVA with association (associated, non-associated) and sentence coherence (coherent, scrambled) as factors revealed, as expected, a significant interaction between association and coherence, $F(1, 118) = 8.894$, $p = .003$, partial $\eta^2 = .070$. It also showed significant main effects of association, $F(1, 118) = 26.170$, $p < .001$, $\eta^2 = .182$, and coherence, $F(1, 118) = 18.564$, $p < .001$, $\eta^2 = .136$. Post-hoc analyses found that responses to associated items were significantly faster than responses to non-associated items after coherent sentences ($t = 5.95$, $p < .001$) but not after scrambled sentences ($t = 1.83$, $p = .071$). Also, responses to items following ordered sentences were significantly faster than responses to items following scrambled sentences for associated items ($t = 5.36$, $p < .001$) but not for non-associated items ($t = 1.13$, $p = .261$). These findings are illustrated in Figure 6.

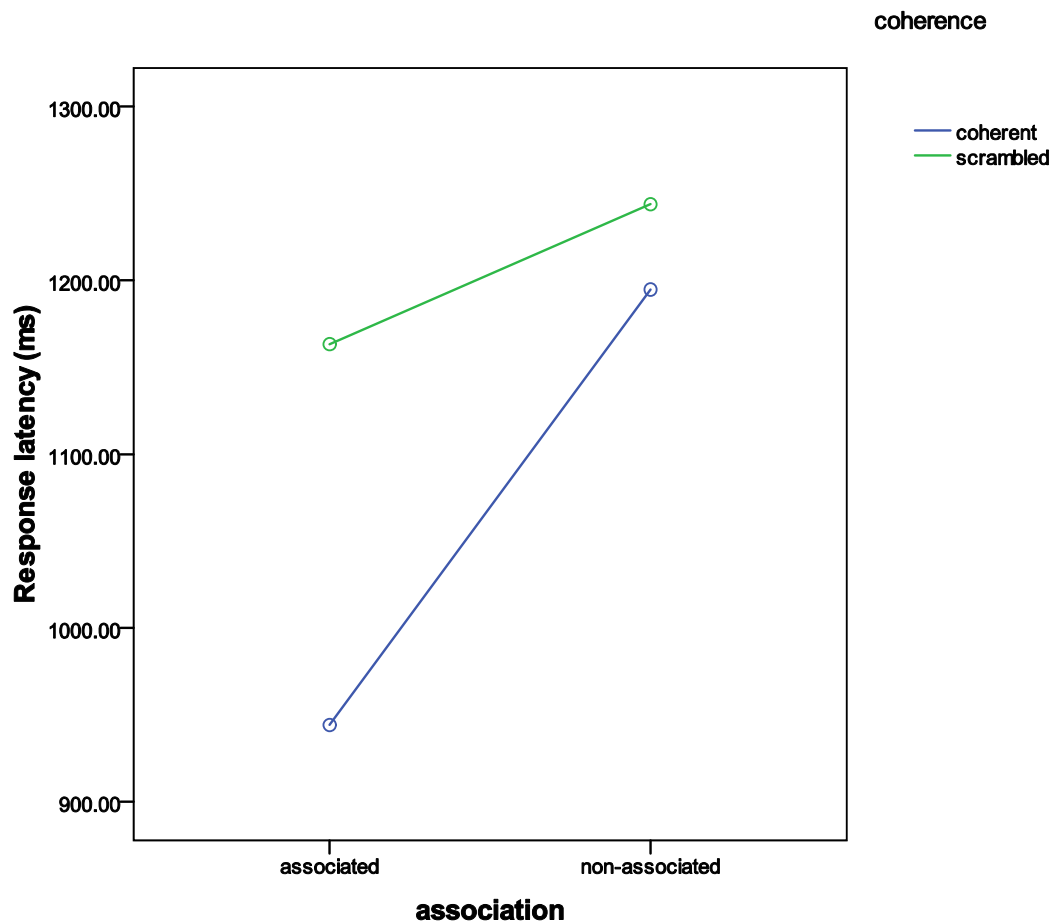


Figure 6: Mean response latencies to words as a function of association and sentence coherence

These results showed that response latencies were shorter when associated words rather than non-associated words followed coherent sentences, but not when associated words rather than non-associated words followed scrambled sentences. Priming of associated words relative to non-associated words, therefore, appeared to result from the generation of inferences rather than from semantic associations between target items and words in the preceding sentences.

Having ascertained that the test appeared to be assessing on-line inference generation across the sample as a whole, the next analysis was conducted to investigate whether there was a relationship between on-line inference generation and performance on the Listening Comprehension Test Series.

Correlational analysis

As explained previously, an on-line inferencing score for each child was calculated using the formula:

$$[RT \text{ (Non-associated/Coherent)} - RT \text{ (Associated/Coherent)}] - [RT \text{ (Non-associated/Scrambled)} - RT \text{ (Associated/Scrambled)}]$$

A Pearson's correlational analysis showed that the on-line inferencing score was not significantly associated with performance on the Listening Comprehension Test Series ($r = .013, p = .893$).

These results were unexpected. It appears that, across the group as a whole, listening comprehension is not associated with on-line inferencing ability. However, it was possible that differences in performance on the test would be evident when only the extreme groups were considered. The next set of analyses, therefore, explored whether or not the pattern of performance on the test differed for the most skilled and least skilled listening comprehenders.

Does the test distinguish between extreme groups of good and poor listening comprehenders?

Children were identified whose listening comprehension score was in either the top third or bottom third of the group of 118. Thirty-seven poorer comprehenders and 38 better comprehenders were identified. The difference in comprehension ability between the groups was significant ($t(73) = 22.436, p < .001$).

To find out whether or not the children who were the most skilled listening comprehenders made knowledge-based inferences on-line whilst the children who were the least skilled did not, a 3-way mixed ANOVA was conducted with association (associated, non-associated) and sentence order (coherent, scrambled) as within-subjects factors and comprehension group (good comprehenders, poor comprehenders) as a between-subjects factor. In the scrambled condition, where inferencing was not possible and priming could result from semantic association only, no difference between good and poor comprehenders was expected. In the coherent

condition, where inferencing was possible, it was expected that good comprehenders but not poor comprehenders might respond faster to associated items. If this occurred, a 3-way interaction between comprehension group, association and coherence would be evident.

Table 26 shows descriptive statistics for each condition for each comprehension group.

Table 26: Means and SD's for response latencies to words in each condition for each comprehension group

	Good comprehenders		Poor comprehenders	
	Associated	Non-associated	Associated	Non-associated
Coherent	951.66 (333.62)	1146.74 (378.45)	922.44 (348.62)	1279.16 (483.32)
Scrambled	1188.62 (482.64)	1210.65 (460.37)	1120.35 (427.96)	1278.72 (498.78)

There was a significant interaction between association and coherence, $F(1, 73) = 6.563$, $p = .012$, partial $\eta^2 = .082$. The main effect of association was significant, $F(1, 73) = 21.151$, $p < .001$, partial $\eta^2 = .225$, as was the main effect of coherence, $F(1, 73) = 10.707$, $p = .002$, partial $\eta^2 = .128$. Post-hoc analyses found that responses to associated items were significantly faster than responses to non-associated items after coherent sentences ($t(74) = 5.13$, $p < .001$) but not after scrambled sentences ($t(74) = 1.63$, $p = .107$). Also, responses to items following coherent sentences were significantly faster than responses to items following scrambled sentences for associated items ($t(74) = 4.45$, $p < .001$) but not for non-associated items ($t(74) = 0.58$, $p = .564$). In other words, across these two groups of comprehender, priming appeared to occur when associated words followed coherent sentences, but not when associated words followed scrambled sentences.

Importantly, the 3-way interaction between comprehension group, association and coherence was found to be non-significant $F(1, 73) = .030$, $p = .862$, partial $\eta^2 = .000$.

It appears, therefore, that the pattern of performance on the test did not differ significantly for the two groups of comprehender. The main effect of group was non-significant, $F(1, 73) = .119$, $p = .731$, partial $\eta^2 = .002$. There were non-significant interactions between association and group, $F(1, 73) = 3.503$, $p = .065$, partial $\eta^2 = .046$, and between coherence and group, $F(1, 73) = 0.461$, $p = .499$, partial $\eta^2 = .006$.

The two groups of comprehender exhibited similar patterns of performance, thus confirming the results of the correlational analysis presented earlier. Performance on the test does not distinguish between good and poor listening comprehenders.

Methodological issues

Results presented here suggested that on-line inferencing ability was not related to listening comprehension in children. However, this conclusion was based on the assumption that the test measured on-line inferencing ability. Although the pattern of results shown in Figure 6 suggested that this was the case, there was a methodological issue with the test which made interpretation of Figure 6 problematic. The fact that materials were not counterbalanced meant that all children saw the same words in the same conditions. Potentially, the faster reaction times to items in the associated/coherent condition, relative to those in the other conditions, could have resulted simply from the materials used. The target words could themselves have elicited faster reaction times. The words appearing in the associated/coherent condition were “birds”, “sick” and “rain”. It was possible that baseline response latencies to these words were simply faster than baseline response latencies to the words in the other conditions. Alternatively, the sentences used in the associated/coherent condition may have contained words which strongly primed the target words regardless of whether or not a topic-related inference was made. For example, the word “rain” followed the sentence “The weather outside was horrible and he looked for his umbrella”. The word “rain” may have been primed by its semantic association to the words “weather” and “umbrella”, rather than by the situation described by the sentence.

The methodological problems with the inferencing test, highlighted here, meant that it could not be concluded that on-line inferencing ability was not related to listening comprehension. The possibility that the results obtained were an artefact of the

materials used had to be investigated. Therefore, a follow-up study was undertaken to investigate the possibility that the relatively fast response times to the words in the associated/coherent condition were due to the materials used and not inferencing *per se*.

7.5. Follow-up study

Summary of Aims

In Study 4, the lack of counterbalancing of materials made interpretation of results difficult. Although it appeared that priming due to inference generation was being detected, the faster response times to words in the associated/coherent condition might have been due to (i) semantic priming of target words by individual sentence words or (ii) faster baseline response latencies to words chosen to be in this condition. A follow-up study was conducted to investigate these possibilities. The study was designed to investigate both the influence of the target items themselves on response latencies, and the extent to which they were primed by their semantic relatedness to words in the preceding sentences.

Due to the choice of materials used in the original study, full counterbalancing across all four conditions in the follow-up study was not possible. Nevertheless, using partial counterbalancing it was possible to investigate whether the faster response latencies to words in the associated/coherent condition compared to those in the associated/scrambled condition were due to inference generation rather than to the stimulus materials used. This was done by comparing response latencies in these two conditions when the same words appeared in each condition. If it could be demonstrated that response latencies to target words were still faster in the associated/coherent condition than in the associated/scrambled condition, the faster response latencies in the associated/coherent condition must be due to inference generation rather than (i) semantic priming of target words by individual sentence words or (ii) differences in baseline response latencies to target words.

It was, therefore, predicted that, when the same materials appeared in both conditions, response latencies in the associated/coherent condition would be faster than response

latencies in the associated/scrambled condition due to the generation of inferences in the associated/coherent condition.

Method

Design

Two versions of the test were used. The first version was exactly the same as that used in the original study. In the second version of the test, associated words which had previously followed coherent sentences now followed scrambled sentences, whilst the associated words which had previously followed scrambled sentences now followed coherent sentences. All other stimuli, including the non-associated words following coherent or scrambled sentences, were exactly the same in both versions of the test. Each child saw one of the versions only. By collapsing across the versions, reaction times to words in the two conditions, associated/coherent and associated/scrambled, could be compared when the same words had appeared in both conditions.

Participants

Fifty-six 8-9 year old children who had not taken part in any of the other studies reported in this thesis were tested. Testing took place in the first term of Year 4.

Materials and procedure

The children were randomly allocated to one of two versions of the test. The first was exactly the same as that used in Study 4. In the second version of the test, associated words which had previously followed coherent sentences now followed scrambled sentences, whilst those previously following scrambled sentences now followed coherent sentences. For example whilst in the first version “Birds” followed “The parents flew back and forth with worms for their babies”, in the second version it followed “The worms flew back and forth with babies for their parents”. Similarly, whilst in the first version, “stuck” followed “He wriggled and could but he just get not wriggled out”, in the second version it followed “He wriggled and wriggled but he just could not get out”. Apart from these differences, the stimuli in the two versions of the test were identical. In this way, response latencies to words in the two conditions, associated/coherent and associated/scrambled, could be compared when the same words had appeared in both conditions. It should be noted that in this follow-up study,

other word items, i.e. those appearing after non-associated sentences, both coherent and scrambled, were treated as filler items only.

Results

Data preparation

One child taking part was excluded from the analysis because they had a statement of Special Educational Needs. Another was excluded due to hearing problems.

Outliers were calculated separately for each of the two conditions. Data points were considered to be outliers if they were greater than 3 SD's from the mean reaction time to words in the condition. Outliers were eliminated, as were incorrect responses. Two children were excluded as the screening procedures meant that they had fewer than 2 correct responses within the acceptable time limits in at least one of the conditions of interest. This left a sample of 52 children.

An *a priori* power analysis was conducted. A medium effect size was assumed based on results obtained in Study 4 ($d = .50$), as was an alpha level of .05 and a beta level of .20. Results suggested that a sample size of 34 was required. The sample size of 52 easily met these minimum requirements.

Are response latencies to associated words faster when they follow coherent sentences than when they follow scrambled sentences?

Response latencies to target words in associated/coherent and associated/scrambled conditions were compared collapsing across lists.

Table 27: Means and SD's for response latencies to words in each condition in follow-up study

<i>Condition</i>	<i>Mean RT</i>	<i>SD</i>
Associated/Coherent	840.83	288.73
Associated/Scrambled	926.50	316.95

A one-tailed paired samples t-test, comparing RTs in associated/coherent and associated/scrambled conditions, showed that the difference between the conditions was significant ($t(51) = 2.01$, $p(\text{one-tailed}) = .025$).

In other words, when the same words appeared in both associated/coherent and associated/scrambled conditions, reaction times to associated words following coherent sentences were faster than reaction times to associated words following scrambled sentences.

In Study 4, fast reaction times were identified to words in the associated/coherent condition relative to the other conditions. Results from the follow-up study suggest that these findings were unlikely to have been due solely to the stimuli used. Even when the same stimuli appeared in both the associated/coherent and associated/scrambled conditions, a difference in reaction times between the two conditions was found. Response latencies to words were faster when they followed coherent sentences and could be primed by the overall context of the sentences than when they followed scrambled sentences and could be primed only by individual words in the sentences. This suggests, therefore, that the test was detecting priming due to the generation of inferences.

7.6. Discussion

The aim of this study was to ascertain whether listening comprehension skill in children was related to their ability to generate inferences on-line. Furthermore, it was hoped that the relationship of on-line inferencing ability with listening comprehension could be explored after controlling for a wide range of other cognitive-linguistic predictors.

The test designed appeared to detect inferencing. There were concerns that, because the materials had not been counterbalanced, target words appearing in the associated/coherent condition may simply have elicited faster response times than words appearing in the other conditions. Furthermore, there were also concerns that inference generation had been confounded with semantic relatedness and that children

were responding most quickly to the words in the inferencing condition simply because of the opportunities that the materials in this condition provided for semantic priming. These possibilities were investigated in a follow-up study, which suggested that the original test was in fact detecting inferencing. Response latencies to associated words were faster when they followed coherent sentences than when they followed scrambled sentences.

Having ascertained that the test detected inferencing, results from the correlational analysis were then unexpected. It appeared that listening comprehension skill was not related to the generation of on-line inferences. Also, when good and poor comprehenders were compared on their pattern of performance on the test, no differences were found.

Whilst performance on the test used in this study was not related to listening comprehension ability, it would be unwise to generalise beyond the materials used here to conclude that good and poor listening comprehenders never differ in on-line inference generation. One limitation of the inferencing test used in this study was the small number of items in each condition. For each of the four conditions in which words were presented, there were only three items. The decision to keep the number of items per condition small was motivated by the need to ensure that the children were able to maintain concentration for the duration of the test. Nevertheless, it should be noted that the small number of items in each condition means that findings may not be reliable.

It may be the case that listening comprehension is not related to the generation of very simple knowledge-based inferences when material is presented in single short sentences. Inferences can be either automatic or strategic (van den Broek, 1994). It could be argued that the inferences generated in this study resulted from automatic processes which did not distinguish between good and poor comprehenders.

It is possible, for example, that the inferencing test tapped primarily contextual semantic priming effects. It has previously been shown that word recognition is facilitated more by contextual priming than by lexical priming (Simpson, Peterson, Casteel, & Burgess, 1989). Simpson et al. compared the time taken to recognise a

word when it followed a coherent sentence and when it followed the same sentence which had been scrambled. Priming effects were much larger after the coherent than after the scrambled sentences despite the fact that the words were the same in both. Furthermore, it has previously been reported that good and poor reading comprehenders differ on semantic priming, based on single word primes, only when semantic associations between words are abstract (Nation & Snowling, 1999). When pairs of words co-occur in the real world and semantic associations between them are concrete, poor comprehenders show as much priming as good comprehenders. It could be speculated, therefore, that in the present study, the faster response latencies to words following coherent rather than scrambled sentences arose from contextual semantic priming effects which tapped automatic processes and were unable to distinguish between good and poor comprehenders.

In the introduction to this chapter, it was argued that it is important to assess children on their on-line inferencing skill. This remains the case, but several issues raised here need to be taken into account when designing a test in future. For example, to control for the fact that different words may elicit different response times from children, it is important that materials are fully counterbalanced. Also, it is important to demonstrate that any priming that occurs is due to inferencing rather than lexical association. It is essential, therefore, that semantic associations between sentence words and target words are weak and few in number. Furthermore, whilst it is necessary to include conditions which allow for semantic association to be controlled for, the inclusion of scrambled sentences may not be the most appropriate method to use. An alternative approach taken by Potts et al. (1988) is the inclusion of coherent sentences which do not describe a situation related to the target word but which contain individual words which are semantically related to the target. The same target also follows coherent sentences containing the same semantically related words which are ordered in such a way that the situation described is related to the target. Response latencies to targets following these two types of sentence are compared so that inference generation can be assessed whilst controlling for any semantic priming of the target that occurs during the processing of coherently structured sentences.

Finally, future investigations should explore the generation of inferences necessary to establish coherence across sentence boundaries, as has been done in the adult

comprehension literature (Long et al., 1994; Potts et al., 1988). With these changes in place, it should be possible to fully explore this important aspect of comprehension in children.

Chapter 8.

General Discussion

8.1. Background to studies

The exact nature of the relationship between reading comprehension and listening comprehension has been debated for decades. Proponents of a “unitary view” of comprehension have argued that listening comprehension and reading comprehension are underpinned by the same processes (Gernsbacher, Varner, & Faust, 1990), whilst others have argued that the different cognitive demands made by comprehension in the two modalities means that, at the very least, different strategies may be useful (Carlisle & Felbinger, 1991; Danks & End, 1987; Rubin, 1980). Even Hoover and Gough (1990), strong proponents of the influential Simple View of Reading, have commented that comprehension in the two modalities may not be exactly the same. Yet, the exact ways in which they differ, in terms of the demands made by comprehension in each modality which are over and above those shared with comprehension in the other modality, have not previously been investigated.

The introduction of the Simple View of Reading into the National Curriculum for England and Wales means that the nature of the relationship between reading comprehension and listening comprehension is of practical, as well as theoretical, importance. Practitioners are now urged to teach listening comprehension skills to children (DfES, 2006). Yet the evidence base specific to listening comprehension is limited. It is simply assumed that listening comprehension involves the same component skills as reading comprehension, with the exception of word recognition ability. The overall aim of the research presented in this thesis was to investigate this assumption.

In this chapter, findings from the empirical studies are interpreted in terms of the research questions that they address and their relationship to existing research evidence. Implications for both theory and practice are then outlined. Finally, limitations of the empirical studies and ideas for future research are discussed.

8.2. Summary of findings

1. Are reading comprehension and listening comprehension predicted by the same cognitive-linguistic variables?

One of the aims of the first two studies was the identification of the most important predictors of reading comprehension and listening comprehension when a range of language, memory and intelligence variables were assessed. As suggested by the Simple View of Reading, it was predicted that the same cognitive-linguistic predictors would be important in comprehension in both modalities, with additional variance in reading comprehension being explained by word recognition skills.

In Study 1, children's comprehension ability was assessed using standardised tests, the NARA-II and the Listening Comprehension Test Series. Concerns were raised regarding differences in these tests, so in Study 2 children's comprehension was assessed using true/false tests of comprehension which utilised similar materials and made similar demands on comprehenders, regardless of modality of presentation. In both studies, reading comprehension was found to be uniquely predicted by word recognition accuracy, vocabulary and syntactic abilities. Both studies found listening comprehension to be uniquely predicted by non-verbal intelligence and vocabulary. However, when listening comprehension was assessed using the standardised test, inferencing ability was also identified as a unique predictor.

Although the results were slightly different in the two studies, they suggested that, whether using standardised or true/false tests, listening comprehension and reading comprehension do not make exactly the same demands on comprehenders. Different abilities appear to underpin reading comprehension and listening comprehension. Whilst vocabulary knowledge is important in performance on all four measures of comprehension used, reading comprehension appears to be more dependent on word- and sentence-level linguistic processes, and listening comprehension is more dependent on discourse-level processes and general cognitive skills.

The finding that inferencing ability did not predict performance on either of the true/false tests of comprehension provides empirical support for the argument made by Cain and Oakhill (2006b) that true/false tests do not adequately assess inferencing ability. The reliability of these tests was potentially problematic, so findings from Study 2 have to be interpreted with care. It was surprising, however, that some of the variables previously identified as predictors of NARA-II reading comprehension, particularly working memory and inferencing, were found to make no unique, significant contributions to NARA-II comprehension in Study 1. Two possible reasons for this, the large number of predictor variables under consideration and the sampling method used, are explained here.

Firstly, with a few exceptions (Cain, Oakhill, & Bryant, 2004; Goff, Pratt, & Ong, 2005), studies of reading comprehension in children have tended to focus on a small number of variables at a time. The problem with this approach is that variables may appear to be significant predictors of reading comprehension simply because other variables have not been controlled for. For example, as in the study reported here, Goff et al. found reading comprehension to be independently predicted by word reading ability, vocabulary skills and syntactic ability but not by performance on a numerical working memory task. Other studies have also identified syntactic ability as a predictor of reading comprehension (Muter, Hulme, & Snowling, 2004) but, surprisingly, this variable has not been controlled for in studies which have demonstrated the importance of inferencing in reading comprehension (Cain et al., 2004).

Secondly, following Goff et al., the current study used an unselected sample of children. This meant that word recognition skill explained a large proportion of the variance in performance on the NARA-II. Most studies of children's reading comprehension have compared children who differ on reading comprehension skill but who are matched on word recognition abilities and on the number of stories on the NARA attempted (e.g. Cain & Oakhill, 1999). Even some of the studies which involve relatively large samples of children have excluded good and poor readers so that the correlations between word recognition and NARA comprehension are low or non-significant (Cain, Oakhill, & Bryant, 2004). In these previous studies, poor reading comprehenders appeared to have poorer inferencing abilities than their more

skilled peers. However, findings from the study reported here suggest that, in an unselected group of readers, inferencing ability does not account for additional variance in reading comprehension once word reading and sentence-level linguistic skills are accounted for.

The analyses reported here did not test the Simple View of Reading directly. This issue was also addressed and findings are summarised in the following section.

2. Is there evidence for the Simple View of Reading?

The main aim of Studies 1 and 2 was to investigate whether there was support for the Simple View of Reading. The question was addressed by investigating whether variance in comprehension in each modality could be systematically explained by any of the cognitive-linguistic predictors after controlling for comprehension in the other modality.

As predicted by the Simple View of Reading, when comprehension was assessed using standardised tests, reading comprehension was positively predicted by word recognition ability only, once listening comprehension was controlled for. All systematic variance in NARA-II reading comprehension could be explained by word recognition ability and performance on the Listening Comprehension Test Series. This provides very good support for the Simple View of Reading and is in agreement with previous research findings (e.g. Byrne & Fielding-Barnsley, 1995; Hoover & Gough, 1990; Neuhaus, Roldan, Boulware-Gooden, & Swank, 2006; Savage, 2001). It is a particularly impressive result given the fact that the materials used to assess reading comprehension and listening comprehension were very different.

Using true/false tests, however, it was found that syntactic skills continued to predict variance in reading comprehension after controlling for listening comprehension, suggesting that the true/false test of reading comprehension makes demands on comprehenders' syntactic skills over and above those made by the true/false test of listening comprehension. This finding supports the argument that the Simple View of Reading does not account for all systematic variance in reading comprehension and that a third variable should be included in the framework. Whilst various possible third variables have been suggested, including processing speed (Joshi & Aaron,

2000) and attentional control (Conners, 2009), the potential role of syntactic skills as a third variable in the prediction of reading comprehension has not previously been explored. Yet, it has been suggested that prosody in speech aids syntactic processing when listening (Rubin, 1980). It might, therefore, be expected that, compared to listening comprehension, reading comprehension should make extra demands on children's syntactic skills. What is unclear is why, in the studies presented in this thesis, this only appeared to be the case when the true/false tests were used. This is an issue that warrants further investigation.

Results regarding the prediction of variance in listening comprehension, having controlled for reading comprehension, were harder to explain using the framework provided by the Simple View of Reading. When standardised tests were used, listening comprehension continued to be positively predicted by vocabulary and inferencing skill once reading comprehension was controlled for. Using true/false tests, listening comprehension was predicted by non-verbal ability only, after controlling for reading comprehension. As might be expected, when more similar materials were used to assess comprehension in the two modalities, vocabulary demands appeared to be shared by listening comprehension and reading comprehension. Although the results differ in the two studies, it is clear that, whichever test is used, listening comprehension makes demands on comprehenders which are beyond those shared with reading comprehension.

This is an interesting finding. The Simple View of Reading provides a framework for understanding reading comprehension not listening comprehension. Gough and Tunmer (1986) do, however, state that "once the printed matter is decoded, the reader applies to the text exactly the same mechanisms which he or she would bring to bear on its spoken equivalent." (p.9). The findings presented here challenge this view. The Simple View of Reading, and indeed the "unitary view" of comprehension, do not account for the systematic prediction of variance in listening comprehension by any other cognitive-linguistic skills after controlling for reading comprehension. The assumption that exactly the same skills are involved in reading comprehension and listening comprehension is not supported. It would appear that, just as reading comprehension makes extra demands on comprehenders' word recognition abilities and, depending on the tests used, on their syntactic skills, so listening comprehension

makes extra demands on listeners above and beyond those shared with reading comprehension.

It is important to emphasise that this does not by itself invalidate the Simple View of Reading as a framework for understanding the component skills involved in reading. Indeed, when standardised tests were used to assess comprehension, strong support was provided for Gough and Tunmer's (1986) assertion that reading comprehension is determined by listening comprehension ability and word recognition skills. Taken together, the results presented here suggest that there is a pool of comprehension skills which are shared by comprehension in the two modalities. However, each modality also makes its own extra demands on comprehenders. Reading comprehension makes demands on word recognition skills and, possibly, on syntactic skills. According to the test used, listening comprehension makes extra demands on vocabulary and inferencing abilities or on non-verbal skills. The implication of these findings for both theory and practice will be discussed later.

3. Is inferencing ability an important predictor of listening comprehension solely because performance on both tasks requires a good memory for explicitly stated content?

The finding that inferencing ability continued to predict variance in performance on a standardised test of listening comprehension, having controlled for reading comprehension was felt to be worthy of further investigation. It was hypothesised that the greater importance of inferencing in listening comprehension that was found in Study 1 may have resulted simply from the shared memory demands of the Listening Comprehension Test Series and the aurally administered test of inferencing. The aim of Study 3 was to ascertain whether or not this was the case.

It was found that good listening comprehenders possessed better memory for explicit material contained in a text than did poorer comprehenders, but that listening comprehension continued to be predicted by the ability to answer questions requiring inference generation even when memory for the text had been controlled for. In other words, the finding that inferencing was an important predictor of listening comprehension was not purely due to the shared memory demands of the two tasks.

Furthermore, memory for explicit information did not make a contribution to listening comprehension that was independent of the contribution of inferencing ability.

These findings are consistent with a large body of research evidence which suggests that good comprehension depends upon the construction of an elaborated, coherent, networked representation of text during encoding which enables access to both explicit and inferred information at test (Long, Oppy, & Seely, 1994; Long, Johns, & Jonathan, (submitted); Oakhill, 1984; Paris & Upton, 1976; van den Broek, Lorch, Linderholm, & Gustafson, 2001).

The findings can also be interpreted in terms of Kintsch's (1988) Construction-Integration model, described in Section 2.3. According to this model, the generation of inferences during comprehension has the effect of strengthening the memory representation of literal text propositions (Perfetti, Landi, & Oakhill, 2005). McNamara, Kintsch, Songer, and Kintsch (1996) argue that questions requiring the generation of inferences interrogate the situation model of a text which has been constructed, whilst questions of memory for explicit material assess the textbase representation. From the results presented here it would appear that good listening comprehenders are those children who are best able to construct a coherent situation model of the text which in turn strengthens their textbase representation.

It may be speculated that the Study 1 finding, that inferencing ability is more important in listening comprehension than in NARA-II reading comprehension, results from the enhanced importance of situation model formation when listening. Anecdotally, it was apparent, when testing the children on the NARA-II, that they frequently sought answers in the text and, in some cases, gave answers which, although marked as correct according to the guidelines, did not seem to reflect a genuine understanding of the material. This strategy of re-visiting the text to find answers is obviously unavailable to listeners. Previous research has also found that children are more likely to generate inferences necessary for text coherence when listening to, rather than reading, text (Hildyard & Olson, 1978). These authors concluded that children adopt different strategies when hearing and reading text. When listening, they argued, children are more focussed on creating a coherent representation of the discourse as a whole.

The possible reasons why children may vary in their ability to construct an on-going situation model of the text are not addressed by the data presented in this thesis. A candidate explanation is that children vary in the “standards of coherence” that they adopt. This has been postulated as an explanation for poor reading comprehension (Cain & Oakhill, 1999; Perfetti, Landi, & Oakhill, 2005) and experimental manipulations that encourage the adoption of high standards of coherence do appear to facilitate inferencing (Hannon & Daneman, 1998) and to improve recall for text (van den Broek et al., 2001).

It must be noted, however, that whilst the data from Study 3 are compatible with the explanation that children generate a situation model on-line, during encoding, there is no direct evidence for this. The data are equally compatible with the view that good listening comprehenders make inferences at test, requiring an accurate representation of the propositional text-base in memory and good off-line inferencing ability to do so. The final study, discussed in the next section, attempted to determine which of these possible explanations was correct.

4. Is there evidence of a relationship between listening comprehension and the on-line generation of inferences?

Study 4 directly addressed the issue of whether or not children’s listening comprehension depended on their tendency to generate inferences on-line as material was being heard.

A test was developed which, despite some methodological concerns, was believed to assess the on-line generation of inferences. However, the tendency to generate these inferences was not related to performance on the Listening Comprehension Test Series. It was speculated that the inferences were generated as the result of contextual semantic priming effects. These automatic processes would not be expected to discriminate between good and poor comprehenders. It may be that, when more complicated inferences are required such as those requiring the establishment of coherence across sentence boundaries, individual differences in inferencing ability would be related to listening comprehension. However, until this is demonstrated

empirically the role of the on-line encoding of inferences in children's listening comprehension, remains a matter of speculation.

8.3. Implications for theory and research

Broadly speaking, the findings presented here provide good support for the Simple View of Reading as a framework for understanding the component skills involved in reading comprehension, particularly when considering the results obtained from the use of standardised tests of reading comprehension and listening comprehension. Results obtained from the use of true/false tests of comprehension suggest that syntactic knowledge should be included as a third component in the framework provided by the Simple View of Reading. This can be explained theoretically by the increased demands on syntactic processing that arise when prosodic cues are unavailable.

A greater challenge to the view that exactly the same processes are involved in listening comprehension and reading comprehension comes from the findings relating to listening comprehension itself. These suggest that listening comprehension makes demands on comprehenders that are not shared with reading comprehension. The assumption that listening comprehension is reading comprehension without word recognition appears not to be the case. These results support the argument made by those authors who claim that different strategies may be used in reading comprehension and listening comprehension simply because the demands of the two tasks are very different (Carlisle & Felbinger, 1991; Danks & End, 1987; Rubin, 1980). These findings also suggest that listening comprehension may actually make more demands on children than reading comprehension and are consistent with previous research findings that suggest that children with good word recognition skills can, in some situations, find the comprehension of written material easier than the comprehension of spoken material (Horowitz & Samuels, 1985; Royer, Kulhavy, Lee, & Peterson, 1986; Royer, Sinatra, & Schumer, 1990).

The findings reported here have important implications for comprehension research. As far as the results from the use of standardised tests are concerned, it appears that reading comprehension can be explained largely by word recognition abilities and

listening comprehension. Compared to reading comprehension, however, listening comprehension makes additional demands on vocabulary and inferencing ability. These results suggest, therefore, that it is more appropriate to investigate children's comprehension skills by examining their listening comprehension than, as traditionally happens, by examining their reading comprehension. Since reading comprehension can be entirely explained by listening comprehension and word recognition skills, as suggested by the Simple View of Reading, a full understanding of the comprehension of spoken material will, necessarily, lead to an understanding of the processes involved in the comprehension of written text. However, the reverse is not the case. Listening comprehension appears to be more demanding than reading comprehension and needs to be investigated as a research topic in its own right.

Reading comprehension is rarely investigated using unselected samples as this potentially confounds reading comprehension with word recognition skill. Certainly, as mentioned earlier, a great deal of the variance in reading comprehension in Study 1 was attributed to word recognition ability. Reading comprehension is usually investigated by comparing children with different comprehension ability but the same word reading accuracy. However, this means that the comprehension skills of children with inadequate word reading ability are overlooked by members of the research community who are interested in exploring higher-level processes in reading comprehension. The study of listening comprehension has the advantage of involving all children, regardless of word recognition ability, without confounding comprehension with word reading skills.

A word of caution should be noted, however. The finding that, when using true/false tests, syntactic skills explain extra variance in reading comprehension having accounted for word recognition abilities and listening comprehension is an interesting one suggesting that, in some respects, reading comprehension may make extra demands compared to listening comprehension. This requires further investigation. Other possible third variables in the Simple View of Reading also need to be explored. Only by conducting research into comprehension in both modalities can a full picture of the skills and processes involved in children's comprehension be obtained.

8.4. Practical implications

It may be expected that the advent of the Primary Literacy Framework in schools will increase the emphasis on listening comprehension, at least in lower primary school settings. The argument put forward in the appendix to the Rose Report makes it quite clear that, when teaching early reading, practitioners need to put considerable effort into developing children's ability to understand oral language (Stuart & Stainthorp, 2006). Of more concern is the issue of whether a focus on listening comprehension is maintained in later years when most children have mastered word reading skills and comprehension tends to be taught through the written modality. Disappointingly, the Listening Comprehension Test Series, demonstrated in this thesis to be an effective test of comprehension, is no longer in publication due to a lack of demand from schools. Yet, evidence presented in this thesis suggests that the assessment and teaching of listening comprehension is important for the following reasons.

Firstly, overall the evidence suggests that reading comprehension is well accounted for by the Simple View of Reading. This means that children with good listening comprehension skills will necessarily have good reading comprehension skills. It is possible that syntactic skills have additional importance in reading comprehension so it is prudent that children have some reading-specific comprehension instruction. Generally speaking, however, it seems that the development of children's listening comprehension skills will benefit comprehension ability in both modalities.

Secondly, it cannot be assumed that children scoring highly on a test of reading comprehension will necessarily possess good listening comprehension skills. As shown, the Listening Comprehension Test Series is a more sensitive test of higher level discourse processes than the NARA-II, suggesting that listening comprehension makes demands on children which are in excess of the demands made by reading comprehension. Children may seem to be competent comprehenders of written text but this may reflect their ability to read words rather than their ability to understand discourse. Because a great deal of time in school is spent listening (Lehto & Anttila, 2003), undiagnosed problems with the comprehension of aurally presented material may affect children's entire school learning experience. It would, therefore appear wise to assess listening comprehension skills directly even in children who appear to have competent reading comprehension skills.

Thirdly, if comprehension is only assessed and taught in the written modality, it is difficult to assess and develop the comprehension abilities of children with poor word recognition skills. It has previously been argued that, whether dyslexic or “garden variety”, it is essential that older poor readers continue to develop their listening comprehension skills as they are unable to benefit from the improvement in comprehension that comes from the experience of reading (Curtis, 1980).

8.5. Limitations and Future Research

Assessment issues

It is important to acknowledge that results presented here are based on the use of certain tests and may not generalise when other measures of comprehension are used. It has previously been reported that predictors of reading comprehension vary according to the tests that are used (Cutting & Scarborough, 2006).

In both Studies 1 and 2, interpretation of results was complicated by the nature of the comprehension tests on which children were assessed. In Study 1, the difference in the materials may have had an influence on the findings whilst in Study 2 there were problems with the reliability and sensitivity of the tests. If further research comparing spoken and written language is to be undertaken, the development of a test which is reliable as a measure of both reading comprehension and listening comprehension is urgently required.

Restricted age-group

In line with previous research in the area, the studies reported here compared reading comprehension and listening comprehension in 7-9 year-old children. It is difficult to know the extent to which the findings generalise to other age-groups. It might be expected that, for older children, with more fluent word recognition skills, the demands of reading comprehension and listening comprehension may be more similar. Certainly, longitudinal studies have suggested that in older age-groups the development of reading comprehension and listening comprehension is highly interdependent (Verhoeven & van Leeuwe, 2008). Further longitudinal studies are

required if the evolving nature of the relationship between reading comprehension, listening comprehension and the predictor variables is to be fully understood.

A further area of future research interest is the relationship between reading comprehension and listening comprehension in adults. The CI model of comprehension (Kintsch, 1998) is thought to apply regardless of modality, an idea supported by Gernsbacher's (1990) notion of a General Comprehension Skill. If it could be demonstrated that differences between comprehension in the two modalities remain in the adult population, existing models and theories of comprehension would require modification.

On-line processing

The case for the assessment of on-line aspects of children's comprehension was made in Study 4. Only on-line tests give information regarding the nature of the processing that occurs during encoding. The interpretation of results from off-line tests often involves speculation regarding the time at which processing is occurring, i.e. during encoding or at test. Yet, on-line aspects of children's processing have barely been addressed in the research literature.

The test used in Study 4, whilst apparently assessing inference generation did not distinguish between good and poor listeners. However, this may be due to the nature of the materials chosen. It is possible that contextual semantic priming effects were in operation. Possible improvements for future studies were suggested. This is a matter not only of theoretical interest but also of practical significance. Whilst it appears, from the studies presented here, that inferencing ability is important in listening comprehension, it is not known whether children actually make inferences whilst they are listening. If this can be shown to be the case, teaching strategies and interventions would need to be focussed on facilitating inferencing during encoding. Within the adult comprehension literature, on-line methods have been used to demonstrate that this can be achieved through embedding questions and slowing presentation of stimuli (Hannon & Daneman, 1998). Similar investigations regarding the facilitation of inferencing in children may be necessary in future.

Listening comprehension – future directions

The results presented here suggest that listening comprehension is an area worthy of investigation in its own right. Future research should examine the relationship of listening comprehension with other variables which were not investigated here but which have also been implicated in reading comprehension, such as comprehension monitoring (Oakhill, Hartt, & Samols, 2005), updating processes (Carretti, Cornoldi, de Beni, & Romano, 2005) and inhibitory deficits (Cain, 2006). Establishing an evidence base concerning the listening comprehension of young children is important, as educational practitioners will need to develop this aspect of children's learning if later reading comprehension is to be successful. Longitudinal investigations are required to clarify the direction of causality when relationships between listening comprehension and predictor variables are apparent. It is not clear, for example, whether, in Study 1, inferencing is necessary for listening comprehension to occur or improves as the result of good comprehension.

It is also necessary to establish an evidence base concerning the effectiveness of any interventions which may facilitate children's listening comprehension development. Within the reading comprehension literature, it has been demonstrated, for example, that comprehension can be improved through the use of inference training activities using written materials (Yuill & Oakhill, 1998; Yuill & Joscelyne, 1998), and through instruction in vocabulary (Beck, Perfetti, & McKeown, 1982). How children can be helped to develop their listening comprehension skills needs to be ascertained.

8.6. Summary/ Original contributions

The research presented in this thesis was undertaken in response to the introduction of the Simple View of Reading into the National Curriculum for England and Wales. This emphasises the importance of children's comprehension of spoken language, yet little research has previously explored this area. The main aim of the work presented here was to investigate the assumption that the same skills and processes underpin reading comprehension and listening comprehension by conducting an investigation of the demands made by comprehension in each modality which are over and above those shared with comprehension in the other modality. This issue has not previously been addressed.

Broad support for the Simple View of Reading was found although it appeared that listening comprehension involved skills and processes which were not shared with reading comprehension. This led to the suggestion that listening comprehension should be treated as a topic worthy of investigation in its own right. In line with this suggestion, two further studies were conducted which explored further the role of inference generation in listening comprehension. Overall it appears that listening comprehension may be related to the tendency to draw inferences during encoding, but this has yet to be confirmed empirically.

References

- Adams, A. M., Bourke, L., & Willis, C. (1999). Working memory and spoken language comprehension in young children. *International Journal of Psychology*, 34 (5/6), 364-373.
- Adams, M. J. (1980). Failures to comprehend and levels of processing in reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer, (Eds.), *Theoretical Issues in Reading Comprehension* (11-32). Hillsdale, NJ: Erlbaum.
- Baddeley, A. (1986). *Working Memory*. Oxford: Oxford University Press.
- Barnes, M. A., Dennis, M., & Haefele-Kalvaitis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology*, 61 (3), 216-241.
- Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, 34 (6), 1373-1399.
- Berger, N. S., & Perfetti, C. A. (1977). Reading skill and memory for spoken and written discourse. *Journal of Reading Behavior*, 9 (1), 7 – 16.
- Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74 (4), 506-521.
- Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology*, 24, 323-335.
- Bishop, D. V. M. (2003). *Test for Reception of Grammar, Version 2*. London: The Psychological Corporation.

Bowey, J. A. (1986). Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology*, 41, 282-299.

Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology*, 75, 189-201.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3, 193-209.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behaviour*, 11 (6), 717-726.

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities*, 40 (3), 226-243.

Burgoyne, K., Kelly, J. M., Whiteley, H. E., & Spooner, A. (2009). The comprehension skills of children learning English as an additional language. *British Journal of Educational Psychology*, 79 (4), 735-747.

Byrne, B., & Fielding-Bransley, R. (1995). Evaluation of a programme to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology*, 87 (3), 488-503.

Cain, K. (2006). Individual differences in children's memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory*, 14 (5), 553-569.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, 11, 489-503.

Cain, K., & Oakhill, J. (2003). Reading comprehension difficulties. In T. Nunes & P. Bryant (Eds.), *Handbook of Children's Literacy* (pp 313-338). Dordrecht: Kluwer Academic Publishers.

Cain, K., & Oakhill, J. (2006a). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, 76, 683-696.

Cain, K., & Oakhill, J. (2006b). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, 76, 697-708.

Cain, K., Oakhill, J., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory and Cognition*, 29 (6), 850-859.

Cain, K., Oakhill, J., & Bryant, P. (2000a). Investigating the causes of reading comprehension failure: The comprehension-age match design *Reading and Writing: An Interdisciplinary Journal*, 12, 31-40.

Cain, K., Oakhill, J., & Bryant, P. (2000b). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing: An Interdisciplinary Journal*, 13, 31-56.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96 (1), 31-42.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: the influence of reading comprehension, vocabulary knowledge and memory capacity. *Journal of Educational Psychology*, 96, 671-681.

Cain, K., Oakhill, J., & Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology*, 90, 65-87.

- Carlisle, J. F., & Felbinger, L. (1991). Profiles of listening and reading comprehension. *Journal of Educational Research*, 84 (6), 345-354.
- Carretti, B., Cornoldi, C., De Beni, R., & Romano, M. (2004). Updating in working memory: A comparison of good and poor comprehenders. *Journal of Experimental Child Psychology*, 91, 45-66.
- Casteel, M. A. (1993). Effects of inference necessity and reading goal on children's inferential generation. *Developmental Psychology*, 29 (2), 346-357.
- Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, 49, 278-293.
- Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities*, 36 (2), 151-164.
- Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending Oral and Written Language* (83-112). London: Academic Press.
- Chen, R. S., & Vellutino, F. R. (1997). Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research*, 29 (1), 1-24.
- Cipielewski, J., & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology*, 54, 74-89.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100 (4), 589-608.

- Conners, F. A. (2009). Attentional control and the simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 22 (5), 591-613.
- Corbett, A. T., & Doshier, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior*, 17 (4), 479-491.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology*, 82 (4), 733-740.
- Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge and spelling. *Journal of Educational Psychology*, 83 (2), 264-274.
- Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, 33 (6), 934-945.
- Curtis, M. E. (1980). Development of components of reading skill. *Journal of Educational Psychology*, 72 (5), 656-669.
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10 (3), 277-299.
- Daneman, M., & Blennerhassett, A. (1984). How to assess the listening comprehension skills of prereaders. *Journal of Educational Psychology*, 76 (6), 1372-1381.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, 3 (4), 422-433.

Danks, J. H., & End, L. J. (1987). Processing strategies for reading and listening. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending Oral and Written Language* (271-292). London: Academic Press.

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, 6 (1), 51-77.

Demont, E., & Gombert, J. E. (1996). Phonological awareness as a predictor of recoding skills and syntactic awareness as a predictor of comprehension skills. *British Journal of Educational Psychology*, 66 (3), 315-332.

DfES (2006). *The new conceptual framework for teaching reading: the “simple view of reading” – overview for literacy leaders and managers in schools and Early Years settings*. London: Department for Education and Skills.

Diakidoy, I-A. N., Stylianou, P., Karefillidou, C., & Papagoergiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology*, 26, 55-80.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale, second edition*. Windsor: NFER-Nelson.

Durrell, D. D. (1969). Listening comprehension versus reading comprehension. *Journal of Reading*, 12 (6), 455-460.

Echols, L. D., West, R. F., Stanovich, K. E., & Zehr, K. S. (1996). Using children's literacy activities to predict growth in verbal cognitive skills: A longitudinal investigation. *Journal of Educational Psychology*, 88 (2), 296-304.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18 (5), 972-992.

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research*, 84 (5), 253-262.

Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology*, 82 (4), 799-804.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.

Florit, E., Roch, M., Altoe, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology*, 27 (4), 935-951.

Freedle, R. O., & Carroll, J. B. (Eds) (1972). *Language Comprehension and the Acquisition of Knowledge*. Washington: V.H. Winston.

Friedman, F., & Rickards, J. P. (1981). Effect of level, review, and sequence of inserted questions on text processing. *Journal of Educational Psychology*, 73 (3), 427-436.

Georgiou, G. K., Das, J. P., & Hayward, D. (2009). Revisiting the "Simple View of Reading" in a group of children with poor reading comprehension. *Journal of Learning Disabilities*, 42 (1), 76-84.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16 (3), 430-445.

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing*, 18, 583-616.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child Psychology*, 63, 563-582.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education*, 7 (1), 6-10.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101 (3), 371-395.

Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing: An Interdisciplinary Journal*, 16, 505-539.

Hagues N., Siddiqui, R., & Merwood, P. (1999). *Listening Comprehension Test Series, Level C*. Windsor: NFER-Nelson.

Hannon, B., & Daneman, M. (1998). Facilitating knowledge-based inferences in less-skilled readers. *Contemporary Educational Psychology*, 23, 149-172.

Healy, J. M. (1982). The enigma of hyperlexia. *Reading Research Quarterly*, 17 (3), 319-333.

Hedrick, W. B., & Cunningham, J. W. (1995). The relationship between wide reading and listening comprehension of written language. *Journal of Reading Behavior*, 27 (3), 425-439.

Hedrick, W. B., & Cunningham, J. W. (2002). Investigating the effect of wide reading on listening comprehension of written language. *Reading Psychology*, 23, 107-126.

Hildyard, A., & Olson, D. R. (1978). Memory and inference in the comprehension of oral and written discourse. *Discourse Processes*, 1 (2), 91-117.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127-160.

Horowitz, R., & Samuels, S. J. (1985). Reading and listening to expository text. *Journal of Reading Behavior*, 17 (3), 185-198.

Horowitz, R., & Samuels, S. J. (1987). Comprehending oral and written language: Critical contrasts for literacy and schooling. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending Oral and Written Language* (1-52). London: Academic Press.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge: Cambridge University Press.

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing*, 19, 339-361.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, 21, 85-97.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80 (4), 437-447.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, 78 (4), 243-255.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99 (1), 122-149.

Kendeou, P., Bohn-Gettler, C., White, M. J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, 31 (3), 259-272.

Kendeou, P., Savage R., & van den Broek, P. (2009). Revisiting the simple view of reading. *British Journal of Educational Psychology*, 79, 353-370.

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, 101 (4), 765-778.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580-602.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95 (2), 163-182.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge: Cambridge University Press.

Kintsch, W., & Rawson, K.A. (2005). Comprehension. In M. J. Snowling and C. Hulme, (Eds.), *The Science of Reading: A Handbook* (209-226). Oxford: Blackwell Publishing.

Kintsch, W., & Kozminsky, E. (1977). Summarizing stories after reading and listening. *Journal of Educational Psychology*, 69 (5), 491-499.

Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85 (5), 363-394.

Kirby, J. R., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy*, 42 (2), 75-82.

Laing, S. P., & Kamhi, A. G. (2002). The Use of Think-Aloud Protocols to Compare Inferencing Abilities in Average and Below-Average Readers. *Journal of Learning Disabilities*, 35 (5), 437-448.

LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 1118-1133.

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness as predictors of early reading ability. *Journal of Experimental Child Psychology*, 58, 88-111.

Lee, L. W., & Wheldall, K. (2009). An examination of the simple view of reading among beginning readers in Malay. *Reading Psychology*, 30 (3), 250-264.

Lehto, J. E., & Anttila, M. (2003). Listening Comprehension in Primary Level Grades Two, Four and Six. *Scandinavian Journal of Educational Research*, 47 (2), 133-143.

Long, D. L., Golding, J. M., & Graesser, A. C. (1992). A test of the on-line status of goal-related inferences. *Journal of Memory and Language*, 31, 634-647.

Long, D. L., Johns, C. L., & Jonathan, E. (submitted). A process view of text representation: Unitization and association.

Long, D. L., Johns, C. L., & Morris, P. E. (2006). Comprehension ability in mature readers. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics (Second Edition)* (797-829). Academic Press.

Long, D. L., & Lea, R. B. (2005). Have we been searching for meaning in all the wrong places? Defining the “search after meaning” principle in comprehension. *Discourse Processes*, 39 (2&3), 279-298.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20 (6), 1456-1470.

Long, D. L., Oppy, B. J., & Seely, M. R. (1997). Individual differences in readers' sentence- and text-level representations. *Journal of Memory and Language*, 36, 129-145.

Long, D. L., & Prat, C. S. (2002). Memory for Star Trek: The role of prior knowledge in recognition revisited. . *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28 (6), 1073-1082.

Long, D. L., & Prat, C. S. (2008). Individual differences in syntactic ambiguity resolution: Readers vary in their use of plausibility information. *Memory & Cognition*, 36 (2), 375-391.

Long, D. L., Seely, M. R., & Oppy, B. J. (1999). The strategic nature of less-skilled readers' suppression problems. *Discourse Processes*, 27 (3), 281-302.

Long, D. L., Wilson, J., Hurley, R., & Prat, C. S. (2006). Assessing text representations with recognition: The interaction of domain knowledge and text coherence. . *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32 (4), 816-827.

McGee, A., & Johnson, H. (2003). The effect of inference training on skilled and less skilled comprehenders. *Educational Psychology*, 23 (1), 49-60.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440-466.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge and levels of understanding in learning from text. *Cognition and Instruction*, 14 (1), 1-43.

McNamara, D. S., & McDaniel, M. A. (2004). Suppressing irrelevant information: Knowledge activation or inhibition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30 (2), 465-482.

Megherbi, H., Seigneure, A., & Ehrlich, M-F. (2006). Reading comprehension in French 1st and 2nd grade children: Contribution of decoding and language comprehension. *European Journal of Psychology of Education*, 21 (2), 135-147.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40 (5), 665-681.

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, 73, 139-158.

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language and Hearing Research*, 47 (1), 199-211.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology*, 72, 549-560.

Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 72, 549-560.

Nation, K., & Snowling, M. J. (1998a). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading. *Child Development*, 69 (4), 996-1012.

Nation, K., & Snowling, M. J. (1998b). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, 39, 85-101.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: evidence from semantic priming. *Cognition*, 70, B1-B13.

Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics*, 21, 229-241.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: broader language skills contribute to the development of reading. *Journal of Research in Reading*, 27 (4), 342-356.

Neale, M. (1997). *Neale Analysis of Reading Ability* (Second Revised British Edition). Windsor: NFER-Nelson.

Neuhaus, G. F., Roldan, L. W., Boulware-Gooden, R., & Swank, P. R. (2006). Parsimonious teaching models: Identifying teachable subskills. *Reading Psychology*, 27, 37-58.

Oakhill, J. (1982). Constructive processes in skilled and less skilled comprehenders' memory for sentences. *British Journal of Psychology*, 73, 13-20.

Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, 54, 31-39.

Oakhill, J., & Cain, K. (2007). Introduction to comprehension development. In K. Cain & J. Oakhill (Eds.), *Children's Comprehension Problems in Oral and Written Language: A Cognitive Perspective* (3-40). New York: Guildford Press.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, 18 (4), 443-468.

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, 18, 657-686.

Oakhill, J. V., & Yuill, N. M. (1986). Pronoun resolution in skilled and less skilled comprehenders: Effects of memory load and inferential complexity. *Language and Speech*, 29, 25-37.

Oakhill, J., Yuill, N., & Parkin, A. (1986). On the nature of the difference between skilled and less-skilled comprehenders. *Journal of Research in Reading*, 9 (2), 80-91.

Omanson, R.C., Warren, W. M., & Trabasso, T. (1978). Goals, inferences, comprehension and recall of stories by children. *Discourse Processes*, 1 (4), 337-354

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98 (3), 554-566.

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story? *Reading and Writing*, 23 (2), 189-208.

Palmer, J., MacLeod C. M., Hunt, E., & Davidson, J. E. (1985). Information processing correlates of reading. *Journal of Memory and Language*, 24 (1), 59-88.

Paris, S. G., & Lindauer, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology*, 8, 217-227.

Paris, S. G., Lindauer, B. K., & Cox, G. L. (1977). The development of inferential comprehension. *Child Development*, 48, 1728-1733.

Paris, S. G., & Upton, L. R. (1976). Children's memory for inferential relationships in prose. *Child Development*, 47, 660-668.

Perfetti, C. A. (1985). *Reading Ability*. New York: Oxford University Press.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly: Journal of Developmental Psychology*, 33 (3), 283-319.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (227-247). Blackwell Publishing.

Perfetti, C. A., Marron, M., & Foltz, P. W. (1996) Sources of comprehension failure: Theoretical perspectives and case studies. In C. Cornoldi & J. Oakhill (Eds.), *Reading Comprehension Difficulties: Processes and Interventions* (137-165). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children*. London: The Psychological Corporation.

Potts, G. R., Keening, J. M., & Golding, J. M. (1988) Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, 27 (4), 399-415.

Raghuathan, T. E., Rosenthal, R., & Rubin, D. B. (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods*, 1 (2), 178-183.

Ransby, M. J., & Swanson, H. L. (2003). Reading comprehension skills of young adults with childhood diagnoses of dyslexia. *Journal of Learning Disabilities, 36* (6), 538-555.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114* (3), 510-532.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Coloured Progressive Matrices*. Oxford: Oxford Psychologists Press.

Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11* (3), 235-257.

Roch, M., & Levorato, M. C. (2009). Simple View of Reading in Down's syndrome: the role of listening comprehension and reading skills. *International Journal of Language & Communication Disorders, 44* (2), 206-223.

Royer, J. M. (2001). Developing reading and listening comprehension tests based on the sentence verification technique (SVT). *Journal of Adolescent and Adult Literacy, 45* (1), 30-41.

Royer, J. M., Hastings, C. M., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. *Journal of Literacy Research, 11* (4), 355 – 363.

Royer, J. M., Kulhavy, R. W., Lee, J. B., & Peterson, S. E. (1986). The sentence verification technique as a measure of listening and reading comprehension. *Educational & Psychological Research, 6* (4), 299-314.

Royer, J. M., Sinatra, G. M., & Schumer, H. (1990). Patterns of individual differences in the development of listening and reading comprehension. *Contemporary Educational Psychology, 15*, 183-196.

Rubin, A. (1980). A theoretical taxonomy of the differences between oral and written language. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical Issues in Reading Comprehension* (411-438). Hillsdale, NJ: Erlbaum.

Sabatini, J. P., Sawaki, Y., Shore, J. R., & Scarborough, H. S. (2010). Relationships among reading skills of adults with low literacy. *Journal of Learning Disabilities*, 43 (2), 122-138.

Savage, R. (2001). The simple view of reading: Some evidence and possible implications. *Educational Psychology in Practice*, 17 (1), 17-33.

Savage, R. (2006). Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. *Scientific Studies of Reading*, 10 (2), 143-164.

Schreiber, P. A. (1987). Prosody and structure in children's syntactic processing. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending Oral and Written Language* (243-270). London: Academic Press.

Schulman, A. I. (1971). Recognition memory for targets from a scanned word list. *British Journal of Psychology*, 62, 335-346.

Seigneuric, A., & Ehrlich, M. F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing*, 18, 617-656.

Seigneuric, A., Ehrlich, M. F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing*, 13, 81-103.

Senechal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, 73 (2), 445-460.

Senechal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of Early Literacy Research (Volume 2)* (173-182), New York: The Guildford Press.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition*, 24 (1-2), 139-168.

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., Fowler, A., Dreyer, L. G., Marchione, K. E., Shaywitz, S. E., & Shaywitz, B. A. (1999). Comprehension and decoding: patterns of association in children with reading difficulties. *Scientific Studies of Reading*, 3 (1), 69-94.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, 60, 973-980.

Simpson, G. B, Peterson, R. R., Casteel, M. A., & Burgess, C. (1989). Lexical and sentence context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15 (1), 88-97.

Singer, M. (1976). Context inferences in the comprehension of sentences. *Canadian Journal of Psychology/Review of Canadian Psychology*, 30 (1), 39-46.

Smiley, S. S., Oakley, D. D., Worthen, D., Campione, J. C., & Brown, A. L. (1977). Recall of thematically relevant material by adolescent good and poor readers as a function of written versus oral presentation. *Journal of Educational Psychology*, 69 (4), 381-387.

Snowling, M. J., & Hulme, C. (Eds.) (2005). *The Science of Reading: A Handbook*. Blackwell Publishing.

Spiro, R. J. (1980). Constructive processes in prose comprehension and recall. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical Issues in Reading Comprehension* (245-278). Hillsdale, NJ: Erlbaum.

Spooner, A. L. R., Baddeley, A. D., & Gathercole, S. E. (2004). Can reading accuracy and comprehension be separated in the Neale Analysis of Reading Ability? *British Journal of Educational Psychology*, 74, 187-204.

Spooner, A. L. R., Gathercole, S. E., & Baddeley, A. D. (2006). Does weak reading comprehension reflect an integration deficit? *Journal of Research in Reading*, 29 (2), 173-193.

Spooner, A. L. R. & Willis, C. (in prep). *On-line inference generation in children's comprehension*.

Spring, C., & French, L. (1990). Identifying children with specific reading disabilities from listening and reading discrepancy scores. *Journal of Learning Disabilities*, 23 (1), 53-58.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behaviour Research Methods, Instruments and Computers*, 31 (1), 137-149.

Stanovich, K. E. (1986) Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21 (4), 360-406.

Stanovich, K. E., & Cunningham, A. E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology*, 85 (2), 211-229.

Stanovich, K. E., Cunningham, A. E., & Feeman, D. J (1984). Intelligence, cognitive skills, and early reading progress. *Reading Research Quarterly*, 19 (3), 278-303.

Stanovich, K. E., & West, R. F. (1989) Exposure to print and orthographic processing. *Reading Research Quarterly*, 24 (4), 402-433.

Sternberg, R., & Powell, W. (1983). Comprehending verbal comprehension. *American Psychologist*, 38, 878-893.

Sticht, T. G. (1972). Learning by listening. In R. O. Freedle & J. B. Carroll (Eds.), *Language Comprehension and the Acquisition of Knowledge* (285-314). Washington: V.H. Winston.

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing*, 4, 245-256.

Stothard, S. E., & Hulme, C. (1995). A comparison of phonological skills in children with reading comprehension difficulties and children with decoding difficulties. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 36 (3), 399-408.

Stothard, S. E., & Hulme, C. (1996). A comparison of reading comprehension and decoding difficulties in children. In C. Cornoldi & J. Oakhill (Eds.), *Reading Comprehension Difficulties: Processes and Interventions* (93-112). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Stuart, M. & Stainthorp, R. (2006). The searchlights model-the case for change. Appendix 1 in *Independent review of the teaching of early reading* (pp. 73-93) (The Rose Review). London: Department for Education and Skills.

Stuart, M., Stainthorp, R., & Snowling, M. (2008). Literacy as a complex activity: deconstructing the simple view of reading. *Literacy*, 42, 59-66.

Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence*, 21, 83-108.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics (Fifth Edition)*. Boston, USA: Pearson.

Tannenbaum K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, 10 (4), 381–398.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory and Cognition*, 16, 283-299.

Tiu, R. D., Jr., Thompson, L. A., & Lewis, B. A. (2003). The role of IQ in a component model of reading. *Journal of Learning Disabilities*, 36 (5), 424-436.

Townsend, D. J., Carrithers, C., & Bever, T. G. (1987). In R. Horowitz & S. J. Samuels (Eds.), *Comprehending Oral and Written Language* (217-242). London: Academic Press.

Tunmer, W. E., & Hoover, W. A. (1992). Cognitive and linguistic factors in learning to read. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading Acquisition*. Hillsdale, NJ: Erlbaum.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154.

van den Broek, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*. London: Academic Press.

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, 29 (8), 1081-1087.

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, 39 (2 & 3), 299 – 316.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

Vellutino, F. R., Tunmer, W. E., Jacard, J. J., & Chen, R. S. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, 11 (1), 3-32.

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, 22, 407-423.

Virtue, S., Parrish, T., & Jung-Beeman, M. (2008). Inferences during story comprehension: Cortical recruitment affected by predictability of events and working memory capacity. *Journal of Cognitive Neuroscience*, 20 (12), 2274-2284.

Wiig, E. H., & Secord, W. (1992). *Test of Word Knowledge*. USA: The Psychological Corporation.

Yuill, N., & Joscelyne, T. (1988). Effect of organizational cues and strategies on good and poor comprehenders' story understanding. *Journal of Educational Psychology*, 80 (2), 152-158.

Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology*, 2 (1), 33–45.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension*. Cambridge: Cambridge University Press.

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*, 80, 351-361.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.

Appendix 1: Title Recognition Test

Title Recognition Test Items and Percentage Recognition for Real Items and Foils

Foils are shown in red.

Horrid Henry's Nits	93.7
Shadows and Showers	2.4
The Worst Witch	56.7
The Cat That Did Magic	8.7
The Magic Finger	52.8
The Duchess's Dilemma	0.8
Danny, Champion of the World	55.9
The Hodgeheg	12.6
Whistling to the Moon	8.7
Maisy Magpie	13.4
The Legend of Spud Murphy	13.4
Burglar Bill	79.5
Captain Underpants and the Attack of the Talking Toilets	45.7
The Faraway Tree	16.5
Catching Raindrops	4.7
The Exciting Story of Archibald Arnold	6.3
The Story of Tracy Beaker	97.6
Give me a Garden	2.4
Tigerlight	3.1
Cliffhanger	33.1
Astrosaurs – The Planet of Peril	17.3
Fast Fergie	6.3
A Bear Called Paddington	66.9
The Badger Bush	3.9
Flat Stanley	43.3
Thank-you, Mrs Rosemary	2.4
Wilberforce the Wizard	8.7
Green Eggs and Ham	51.2
Crackers and Cheese	8.7
Ms Wiz Spells Trouble	12.6
The Owl Who Was Afraid of the Dark	66.1
Mrs. Munroe's Muffins	6.3
Big Sister and the Sneeze	3.1
Elephants Can't Tiptoe	6.3
The Vicious Vikings	56.7
Trevor, the Troublesome Toad	3.9
Asterix the Gaul	5.5
Magic Max and the Melting Mystery	6.3
The Iron Man	78.7
Dreaming Doris	6.3

Appendix 2: Mean Responses to TOWK Items

Mean responses to items on “Word Definitions” subscale of the Test of Word Knowledge (Min = 0.00, Max = 2.00)

1. Magician	1.67
2. Envelope	1.58
3. Teacher	1.51
4. Scarf	1.38
5. Bus	1.75
6. Friend	1.34
7. Broom	1.25
8. Mansion	0.90
9. Butterfly	1.28
10. Apple	1.13
11. Bird	1.16
12. Aquarium	0.71
13. Tree	1.42
14. Nephew	0.68
15. Lunch	0.88
16. Architect	0.07
17. Tournament	0.48
18. Chaperone	0.09
19. Jug	0.98
20. Bib	0.95
21. Atlas	0.37
22. Predator	0.37
23. Javelin	0.68
24. Campus	0.01
25. Fragment	0.04
26. Talon	0.09
27. Gavel	0.00
28. Cartilage	0.03
29. Breadwinner	0.00
30. Ember	0.02
31. Poultry	0.02
32. Haven	0.00

Appendix 3: Off-line inferencing test material

Tim and the Biscuit Tin

Tim waited until he was alone in the house. The only sound he could hear was his father's axe on the logs in the shed. Tim looked in all the rooms again, to make sure his mother was not there. Then he pushed a chair over to the sink which was full of dishes. By climbing onto the edge of the sink, he could just reach the biscuit tin. The tin was behind the sugar. Tim stretched until his fingers could lift up the lid. Just as he reached inside, the door swung open and there stood his little sister.

1. Why did Tim want to be alone in the house?

So he could steal/get/eat the biscuits

So he wouldn't be caught/told off

2. What room was Tim in?

Kitchen

3. What was Tim's father doing?

Chopping/cutting logs/wood

4. How did Tim climb onto the sink?

By pushing a chair over to the sink and standing on it

By using/standing on a chair

John's Big Test

John had got up early to learn his spellings. He was very tired and decided to take a break. When he opened his eyes again the first thing he noticed was the clock on the chair. It was an hour later and nearly time for school. He picked up his two books and put them in a bag. He started pedalling to school as fast as he could. However, John ran over some broken bottles and had to walk the rest of the way. By the time he had crossed the bridge and arrived at class, the test was over.

1. How did John travel to school?

By bike/ by bike and walking

2. What did John do when he decided to take a break?

He went to sleep/ had a nap

3. Why did John have to walk some of the way to school?

He got a puncture

Tyre/bike was broken

4. How do you know that John was late for school?

Missed test/ test was over

Linda Does a Kind Thing

Linda was playing with her new doll in front of the house. Suddenly, she heard a strange noise coming from under the bushes. It was the flapping of wings. Tears came to Linda's eyes because she did not know what to do. She ran inside and got a shoe box from the cupboard. Then Linda looked inside her desk until she found eight sheets of yellow paper and some scissors. When she had finished she put the little pieces of paper in the box. Linda gently picked up the helpless creature and took it with her. Her teacher knew what to do.

1. Where did Linda go at the end of the story?

School

2. What creature was making the noise?

Bird/bat/specific type of bird

3. What did Linda do with the paper before she put it in the box?

Cut it up

4. How did Linda feel when she saw what was making the noise?

Upset/sad/tearful

Bill and the Captain go Fishing

The waves were high and the wind was blowing hard. Bill held the edge of his seat with his cold, wet fingers. His tummy felt as if it was going up and down and from side to side at the same time. Bill had been throwing his fishing line towards the beach but now it lay at his feet. Bill watched the smoke drift past. The captain was trying to light his pipe but, each time, the wind blew out the tiny flame. The captain tried once more then threw the empty cardboard box into the bucket. Bill just wanted to go home but, for now, all he could do was hang his head over the side and wait.

1. Why did Bill want to go home?

He felt seasick/ didn't feel well

Tummy felt weird

2. Where was the fishing boat?

On the sea/ocean

3. What had been in the cardboard box?

Matches

4. Where was the smoke coming from?

The Captain's pipe

The thing the Captain was smoking

The matches being blown out/ tiny flame

Appendix 4: True/False Listening Comprehension Test

Practice Story

My friend and I made a tree-house. We like to hide in it. We climb up the rope and pull it up after us. Then no-one knows where we are. We play space-ships. At tea-time we slide down fast and we are always first for tea.

True or false?

- | | |
|---|-------|
| Tree-house would be a good name for that story. | true |
| The boy's dad made the house in the tree. | false |
| They get into the tree-house by climbing up a rope. | true |
| The children play table tennis in the tree-house. | false |

Story 1

I lost my boat. The wind took it out to sea. Then a lost dog went for it. Now I have my boat and a pet dog too.

True or false?

- | | |
|------------------------------------|-------|
| The boy's boat got taken by a dog. | false |
| The dog went to save the boat. | true |
| The boy didn't get his boat back. | false |
| The dog now lives with the boy. | true |

Story 2

Tony and Susan awoke suddenly. The dog was barking loudly in the yard. The children ran to the window. They could see smoke and flashing lights some way off. A helicopter had crash-landed in the park nearby. Flames shot into the air. They saw the pilot jump clear and run to safety.

True or false?

- | | |
|--|-------|
| Tony and Susan were woken up by the dog running into their room. | false |
| They ran to the yard. | false |
| They saw lights and smoke. | true |
| A helicopter had exploded. | false |
| The crash happened in the park. | true |
| The helicopter was on fire. | true |
| The pilot was in danger because he was badly hurt. | false |
| The pilot jumped out of the helicopter. | true |

Story 3

Emma had never been camping before. She woke up in the dark, and heard snorting and rustling outside. As Emma listened, the noise came closer. She grabbed her sister and told her a huge monster was coming. But when her sister looked, she laughed. She told Emma to come and see. It was a hedgehog, hunting for slugs. Emma was amazed that such a small creature could make so much noise.

True or false?

- | | |
|-------------------------------------|-------|
| This was Emma's first time camping. | true |
| The story happened in the day. | false |

Emma was scared by a noise.	true
Emma shared a tent with her sister.	true
Emma's sister laughed when she saw the hedgehog.	true
Emma saw something scary.	false
The hedgehog was hunting for worms.	false
The hedgehog was really big.	false

Story 4

The wizard peered into his crystal ball. Amongst the swirling colours, he could see two children out in the forest. They looked cold and scared. Sarah and Jim stared bleakly at the trees. The path they had followed had vanished and they wondered which way to go. Sarah shivered, and started to sob. Then Jim cupped his ear and whispered excitedly "Listen!" The rhythm of a horse's hooves was coming through the trees. Suddenly, as if from the very darkness, the wizard appeared and took the surprised and grateful children to their village.

True or false?

Sarah and Jim lived in the forest.	false
The children were lost.	true
A path had led them into the middle of the forest and then stopped.	true
The wizard knew that the children were in the forest when he heard Jim speaking.	false
Sarah was particularly upset.	true
Jim heard the hooves of a horse.	true
The children were afraid of the wizard.	false
The wizard told the children how to get out of the forest.	false

Story 5

What excitement to be chosen for a sailing expedition around the world, commemorating the journey of Francis Drake some four hundred years ago! The young explorers had been selected from different nations for their enthusiasm and range of abilities. The imagination of everyone was stirred. During the long voyage the crew would pursue scientific projects and provide community services. Their achievements outstripped the dreams of the explorers and their sponsors. Under the direction of scientists the young people salvaged ancient wrecks, rebuilt houses, mapped jungle trails and used aerial walkways to study tall forests. Some overcame physical disabilities to assist in relief work for an area stricken by a hurricane. Their exploits suggest that courage, adaptability and the spirit of adventure still flourish.

True or false?

The people in the story recreated the voyage of Francis Drake.	true
The two main aims of the voyage were to carry out scientific projects and provide community services.	true
The explorers were selected for their scientific abilities.	false
You know the expedition was a great success because they got all the way round the world.	false
Scientists helped them to carry out projects.	true
They rebuilt ancient wrecks and studied a hurricane.	false
Some of the young explorers had a greater test of their courage than others because their houses got destroyed in the hurricane.	false

In general, the young people showed that they were brave and adventurous.

true

Story 6

Realising the necessity to conserve the strength of the team, the leader decided to pitch an intermediate camp. The initial enthusiasm and anticipation of attaining the final camp had been subdued by the recent mishap in which one member had fallen into a crevasse. Although the rescue had been accomplished magnificently, it was obvious that the incident had hampered the original programme. The team accepted the leader's decision with relief. The tedious crawl to the plateau against incessant winds of varying violence had challenged their endurance to the limit. Every step at this height required will-power. Immediately ahead lay an unforeseen rise from which, by great misfortune, all the tracks of the advance party had disappeared. Rest was essential if the team were to withstand the arduous conditions in the concluding stages of the assault upon this unconquered peak.

True or false?

The leader realised that the team needed rest.

true

The leader decided to set up camp.

true

The team were disappointed with the leader's decision to stop climbing.

false

Their progress had been hampered by one of the team falling into a crevasse.

true

The steep slope had made them slacken their pace of climbing to a crawl.

false

Just ahead of them lay a deep crevasse.

false

It was a piece of bad luck that the tracks of the other party had disappeared.

true

It would be really exciting to reach the peak because then they could rest.

False

Appendix 5: True/False Reading Comprehension Test

Story 1

Mother gave me a big box. I put it on the table. I looked in the box for a toy. Then out jumped a white rabbit.

True or false?

Mother gave the girl a toy.	false
The girl put the box on the table.	true
The girl thought there was a rabbit in the box.	false
The rabbit was a surprise.	true

Story 2

John and Ann were looking for tadpoles. Suddenly they heard a splash. A fisherman has fallen into the lake. He could not swim because he was hurt. The children tried to pull him ashore. He was too heavy. Then Ann held the man's head above water while John raced for help.

True or false?

John and Ann were looking for tadpoles at the lake.	true
The children heard a shout for help.	false
A man fell into the lake.	true
The man could not swim because he was too heavy.	false
The children tried to pull him to the shore.	true
They could not get him ashore because he was hurt.	false
Ann shouted for help.	false
John ran for help.	true

Story 3

As soon as they reached the house, Jack jumped out of the fire-engine and ran inside. He knew a boy was trapped upstairs. He battled through flames to climb the stairs and found Tim hiding behind a bed. Then they heard a loud crash. The stairs had vanished! Jack opened the window and shouted for a ladder. When they reached the ground Tim hugged his mother and Jack was a hero.

True or false?

Jack is a policeman.	false
Jack jumped out of a fire-engine.	true
The boy was trapped downstairs.	false
Tim was hiding in the bathroom.	false
The stairs collapsed.	true
Jack shouted for a hose.	false
They escaped down a ladder.	true
Tim hugged his mother.	true

Story 4

Dark clouds blotted out the fading daylight. A mournful wailing filtered through the deserted building. The children stopped exploring. "Ghosts!" whispered one child. "Nonsense!" replied the other. Nevertheless, they proceeded cautiously in the

direction of the mysterious noise. Gathering courage, and with mounting curiosity, they approached the old kitchen door. Scarcely daring to breathe, they released the catch. Their torches searched the darkness. Immediately their anxiety turned to pity. An exhausted dog lay crouched and whimpering. A gust of wind had slammed the door shut while the dog had been hunting for rats.

True or false?

Dark clouds made it difficult for the children to see where they were going.	true
The children had come into the building to look for ghosts.	false
They stopped suddenly because it went so dark.	false
Then they went to find out what the noise was.	true
The noise was coming from upstairs.	false
They discovered lots of rats.	false
The dog had been hunting for rats.	true
The children felt sorry for the dog because he lay crouched and whimpering.	true

Story 5

The stricken submarine lay at a depth of approximately thirty metres. Although it was common knowledge that the treacherous currents of the area would make rescue operations difficult the crew remained disciplined and confident. Meanwhile, outside their prison, a diver with technical equipment for their release was in peril. His lifeline had become entangled around a projection on nearby wreckage. Experience warned him against his first impulse to dislodge the line by force. Patiently he turned and twisted. At last his calmness and persistence were rewarded. Triumphant he dislodged the final loop from the obstruction. Then weary but undaunted by this unpleasant accident he proceeded to provide an escape exit for the submarine's captives.

True or false?

The diver had gone into the sea to detach an entangled lifeline.	false
He had to go down to a depth of about thirty metres.	true
This part of the sea was noted for its dangerous currents.	true
The crew felt scared.	false
The diver's lifeline got caught on some wreckage.	true
His experience warned him not to twist and turn.	false
The diver was calm and persistent.	true
As soon as he was free he went to get some help.	false

Story 6

The scientist approached the crater's edge fascinated at the prospect of recording the spectacle of a dormant volcano smouldering again. Intent on their photography they ignored an ominous rumbling. Within seconds, the subterranean cauldron exploded violently, ejecting a great quantity of rocks. Fortunately these fell in the direction of the opposite slope. Greatly alarmed by this premature explosion, the group hastily began the descent. Immediately, fiery boulders from a gigantic avalanche hurtled around them. Aware that their apparatus hindered progress, they abandoned all equipment except their precious cameras. Then came an anxious moment. As they were evading flying fragments, one of them was struck off-balance by a rebounding

boulder. A lengthy halt would have been disastrous. Everyone was, therefore, immensely relieved when they discovered the injuries were superficial. They resumed their hazardous scramble to regain safety just before the surroundings were destroyed.

True or false?

The scientists had come to the volcano to record it erupting.	false
The volcano was so interesting because it had been smouldering for many years.	false
The scientists ignored the first explosion.	false
When the volcano erupted the scientists were running down the slope.	false
The first explosion wasn't so dangerous because the rocks fell on the opposite slope.	true
To speed up their decent they dropped their equipment.	true
There was an accident when one person was struck by a boulder.	true
It was vital for them to descend quickly.	true

Appendix 6: Off-line inferencing and memory questions

For stories, please see Appendix 3.

Tim and the Biscuit Tin

1. Why did Tim want to be alone in the house? (Inference) *So he could steal/get/eat the biscuits. So he wouldn't be caught/told off*
2. Where was Tim's father? (Memory) *In the shed*
3. What was Tim's father doing? (Inference) *Chopping/cutting logs/wood*
4. What was in the sink? (Memory) *Dishes/bowls/washing-up*
5. How did Tim climb onto the sink? (Inference) *By pushing a chair over to the sink and standing on it. By using/standing on a chair*
6. Where was the biscuit tin? (Memory) *Behind the sugar*
7. What room was Tim in? (Inference) *Kitchen*
8. What happened as Tim reached inside the tin? (Memory) *His little sister came in/door swung open*

John's Big Test

1. What was John trying to learn? (Memory) *Spellings*
2. What did John do when he decided to take a break? (Inference) *He went to sleep/had a nap*
3. Where was the clock? (Memory) *On the chair*
4. How did John travel to school? (Inference) *By bike. By bike and walking*
5. How many books did John pick up? (Memory) *Two*
6. Why did John have to walk some of the way to school? (Inference) *He got a puncture. Tyre was broken*
7. What did John have to cross on his way to school? (Memory) *Bridge*
8. How do you know that John was late for school? (Inference) *Missed test/ test was over*

Linda Does a Kind Thing

1. What was Linda doing before she heard the strange noise? (Memory) *Playing with doll/toy. Playing in front of house/ outside*
2. What creature was making the noise? (Inference) *Bird/ specific type of bird*
3. Where was the noise coming from? (Memory) *Bushes*
4. How did Linda feel when she saw what was making the noise? (Inference) *Upset/sad/tearful*
5. What did Linda get from the cupboard? (Memory) *(Shoe)box*
6. What did Linda do with the paper before she put it in the box? (Inference) *Cut it up*
7. What colour was the paper? (Memory) *Yellow*
8. Where did Linda go at the end of the story? (Inference) *School*

Bill and the Captain go Fishing

1. Where was the fishing boat? (Inference) *On the sea /ocean*
2. What were Bill's fingers like? (Memory) *Cold and wet/ Cold/ Wet (or synonym)*
3. Why did Bill want to go home? (Inference) *He felt seasick/ didn't feel well
Tummy felt weird*
4. Where was Bill's fishing rod? (Memory) *At his feet. Next to/under his feet*
5. Where was the smoke coming from? (Inference) *The Captain's pipe. The thing the Captain was smoking. The matches being blown out/ tiny flame*
6. What was the captain trying to do? (Memory) *Light his pipe*
7. What had been in the cardboard box? (Inference) *Matches*
8. Where did the captain throw the cardboard box? (Memory) *In the bucket*

Appendix 7: On-line inferencing stimulus materials

The sentences with a word as the target are highlighted in red.

Word type	Coherent/ scrambled	Sentence	Letter string
Verification	Coherent	There was so much to do that mum had not sat down all morning.	Right order?
Non-word	Coherent	You can stroke the animals and walk in the fields.	lote
Word non-associated	Coherent	In the evening they all sat and ate at the table together.	race
Non-word	Scrambled	Even if she couldn't reach she jumped the cupboard.	prale
Word non-associated	Scrambled	They were him taking the police and now the money had caught here.	dark
Non-word	Scrambled	When the match won the ball we knew he had hit the net.	lort
Non-word	Coherent	He loved Christmas and wondered what he'd get.	slad
Word associated	Scrambled	He'd put a long ambulance and the siren fallen its way on.	hurt
Word associated	Coherent	The weather outside was horrible and he looked for his umbrella.	rain
Non-word	Scrambled	There was a loud floor and the crash was all over the glass.	skun
Word associated	Coherent	Mum took his temperature and kept him off school for a week.	sick
Verification	Coherent	He had a uniform and had fought in many wars.	Right order?
Word non-associated	Scrambled	I stopped too fast and he was drove when he trembling.	tree
Verification	Scrambled	Her swings played brother there and there were football for her.	Right order?
Word non-associated	Coherent	She'd got the time wrong and no-one else was there yet.	milk
Non-word	Coherent	They'd had a lovely week and didn't want to go home.	noor
Verification	Scrambled	They see on the miles and could ran for sand.	Right order?
Word associated	Scrambled	He wriggled and could but he just get not wriggled out.	stuck
Verification	Coherent	He watched her walk down the aisle in her beautiful dress.	Right order?
Word associated	Coherent	The parents flew back and forth with worms for their babies.	birds
Non-word	Coherent	All my friends came. We had balloons and cake and played loads of games.	shate

Verification	Scrambled	He hoped up his little tiny sister and picked she wouldn't scream.	Right order?
Non-word	Scrambled	As she was onto the train the platform ran pulling away.	gop
Non-word	Coherent	On his first day he'd already made a friend and was learning to spell.	wask
Word non-associated	Coherent	As the play started he hoped he'd remember all his lines.	lost
Word non-associated	Scrambled	She rubbed very well and she hadn't slept her eyes.	hot
Verification	Coherent	He glanced at the wall to see if it was time to go.	Right order?
Word associated	Scrambled	He took his carpet and dropped ages to clean the plate.	food
Non-word	Coherent	Before his lessons, he'd been scared of the water but now he could join the others.	saud
Non-word	Scrambled	He'd once been a long time but that was a very famous footballer ago.	shog
Non-word	Scrambled	She didn't move her best friend to want away and she cried.	saul
Verification	Scrambled	Mum brought hungry and dad woke up a food of tray.	Right order?